

Università degli Studi di Siena



Facoltà di Scienze Matematiche Fisiche e Naturali

Tesi di Dottorato in Fisica Sperimentale

PhD Thesis in Experimental Physics

XX Ciclo

**Study of the Heavy Flavour fractions in Z +jets events
from $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV with the CDF II
detector at the Tevatron collider**

Candidate

Paolo Mastrandrea

Internal Supervisor

Prof. Angelo Scribano

Supervisor

Prof. Carlo Dionisi

Supervisor

Dr. Marco Rescigno

Contents

Introduction	1
1 Theoretical review	5
1.1 Z boson production at hadron colliders	5
1.1.1 Z boson decay	8
1.2 Associated production of a Z boson and Heavy Flavours	9
1.3 Jet physics	11
1.3.1 The cone algorithm	13
1.4 Monte Carlo generators	14
1.4.1 Alpgen	16
2 The Tevatron and the CDF Detector	19
2.1 The Tevatron	21
2.1.1 The proton beam	21
2.1.2 Main Injector	23
2.1.3 Antiproton Production and Recycler	25
2.1.4 Tevatron	25
2.2 The CDF Detector in Run II	27
2.2.1 System Reference	27
2.2.2 The Tracking System	29
2.2.3 Time of Flight	33
2.2.4 The Solenoid	33
2.2.5 CPR2: the Central Preshower system	34
2.2.6 Calorimeters	34

2.2.7	Muon Detectors	39
2.2.8	Cherenkov Luminosity Counter: CLC	42
2.2.9	Forward Detectors	42
2.2.10	Trigger System	44
2.2.11	Online Monitoring	46
2.2.12	Data Processing	46
3	NN-based algorithm for HF jets identification	47
3.1	<i>b</i> -tagging at CDF II	47
3.2	Neural Network tagger	49
3.2.1	Training samples	51
3.2.2	Vertexing algorithm	52
3.2.3	Vertexes classification	53
3.2.4	Unvertexed tracks	56
3.2.5	Jets discrimination	59
3.2.6	Combined Networks	60
3.3	Data - Monte Carlo comparison	64
4	Calibration procedure	69
4.1	Calibration strategy	69
4.2	Definition of calibration samples	70
4.3	Extraction of the correction functions	74
4.4	Uncertainty of the correction functions	76
5	Z+jets events selection	81
5.1	Data and Monte Carlo samples	81
5.2	Electron selection	85
5.2.1	Electron trigger efficiency	90
5.2.2	Electron selection efficiency	91
5.2.3	$Z \rightarrow e^+e^-$ events selection	93
5.2.4	Electron energy scale	94
5.2.5	Background to $Z \rightarrow e^+e^-$ events	97

5.3	Muon selection	100
5.3.1	Muon trigger efficiency	103
5.3.2	Muon selection efficiency	107
5.3.3	$Z \rightarrow \mu^+ \mu^-$ events selection	107
5.3.4	Background to $Z \rightarrow \mu^+ \mu^-$ events	109
5.4	Z boson selection results	110
5.5	<i>Anti-top</i> cuts	110
5.6	$Z + jet$ events selection	114
5.6.1	Jet corrections	114
5.6.2	Jet selection	117
6	Heavy Flavour fraction measurement	121
6.1	Neural Network distribution for $Z+jets$ events	122
6.1.1	Identification Efficiency	122
6.1.2	Neural Network Monte Carlo templates	124
6.1.3	Correction functions for $Z+jets$ events	125
6.2	Heavy Flavour fraction fit	127
6.3	Measurement of the cross section	131
6.3.1	Identification efficiency correction	132
6.3.2	Detector unfolding	133
6.3.3	Results	135
6.4	Systematic uncertainties	136
6.4.1	Jet energy scale	138
6.4.2	Correction function uncertainties	138
6.4.3	Uncertainty on templates	140
6.4.4	Background normalization	140
6.4.5	Identification efficiency uncertainty	142
6.4.6	Jet unfolding uncertainty	142
6.4.7	Summary of systematic uncertainties	143
6.5	Results and future perspectives	143
A	Fraction fits	145

Bibliography	146
--------------	-----

Introduction

The Standard Model of field and particles is the theory that provides the best description of the known phenomenology of the particle physics up to now.

Data collected in the last years, mainly by the experiments at the big particle accelerators (SPS, LEP, TEVATRON, HERA, SLAC), allowed to test the agreement between measurements and theoretical calculations with a precision of $10^{-3} \div 10^{-4}$.

The Standard Model is a Quantum Field Theory based on the gauge symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$, with spontaneous symmetry breaking. This gauge group includes the color symmetry group of the strong interaction, $SU(3)_C$, and the symmetry group of the electroweak interactions, $SU(2)_L \times U(1)_Y$. The formulation of the Standard Model as a gauge theory guarantees its renormalizability, but forbids explicit mass terms for fermions and gauge bosons. The masses of the particles are generated in a gauge-invariant way by the *Higgs Mechanism via* a spontaneous breaking of the electroweak symmetry. This mechanism also implies the presence of a massive scalar particle in the mass spectrum of the theory, the *Higgs boson*. This particle is the only one, among the basic elements for the minimal formulation of the Standard Model, to have not been confirmed by the experiments yet.

For this reason in the last years the scientific community has been focusing an increasing fraction of its efforts on the search of the Higgs boson.

The mass of the Higgs boson is a free parameter of the Standard Model, but the unitarity of the theory requires values not higher than 1 TeV and the LEP experiments excluded values smaller than 115 GeV. To explore this range of masses is under construction at CERN the Large Hadron Collider (LHC), a proton-proton collider with a center of mass energy of 14 TeV and a $10^{34} cm^{-2} s^{-1}$ peak luminosity. According to the present schedule, this machine will start to provide collisions for the experiments at the end of 2008. In the meanwhile the only running accelerator able to provide collisions suitable for the search of the Higgs boson is the Tevatron at Fermilab, a proton-antiproton collider with a center of mass energy of 1.96 TeV working at $3 \cdot 10^{32} cm^{-2} s^{-1}$ peak luminosity. These features make the Tevatron able for the direct search of the Higgs boson in the 115-200 GeV mass range. Since the coupling of the Higgs boson is proportional to the masses of the particles involved, the decay in $b\bar{b}$ has the largest branching ratio for Higgs mass < 135 GeV and thus the events $Z/W + b\bar{b}$ are the main background to the Higgs signal in the most range favored by Standard Model fits.

In this thesis a new technique to identify Heavy Flavour quarks inside $high - P_T$ jets is applied to events with a reconstructed Z boson to provide a measurement of the Z+b and Z+c inclusive cross sections. The study of these channels represent also a test of QCD in high transferred momentum regime, and can provide information on proton *pdf*.

This new Heavy Flavour identification technique (*tagger*) provides an increased statistical separation between b, c and light flavours, using a new vertexing algorithm and a chain of artificial Neural Networks to exploit as much information as possible in each event.

For this work I collaborated with the Università di Roma “La Sapienza” group working in the CDF II experiment at Tevatron, that has at first developed this tagger.

After a brief theoretical introduction (chapter 1) and a description of the

experimental apparatus (chapter 2), the tagger itself and its calibration procedure are described in chapter 3 and 4. The chapter 5 is dedicated to the event selection and the chapter 6 contains the results of the measurement and the study of the systematic errors.

Chapter 1

Theoretical review

1.1 Z boson production at hadron colliders

At a hadron collider, like the Tevatron at Fermilab, the Z boson is produced in high energy $p - \bar{p}$ collisions. Like all hadrons, p and \bar{p} are bound states of quark and gluons (partons), glued together by the strong interaction. The part of the Standard Model that describes the dynamics of the elementary particles subjected to the strong interaction is called Quantum Chromodynamics.

At the typical energy scale of the hadron colliders, the interaction processes are described by perturbative QCD. However, in order to predict cross sections for such processes the structure of colliding hadrons has to be described in terms of partons.

The parton model relies on the assumption that the low energy dynamics responsible for the parton confinement inside the hadron decouples from the high energy interactions the parton undergoes with any external probe. The time scale of the gluon exchange inside the hadron is in fact of the order of $1/m_h$ (where m_h is the hadron mass), thus any external probe, hitting the hadron with $Q^2 \gg m_h^2$, (where Q is the momentum exchanged in the interaction) will interact with one parton as if it were a free particle. The specific state of such a parton cannot be predicted by perturbative QCD, but, using for instance virtual photons, it is possible to measure the *parton distribution functions* (*pdf*) $f(x_a, Q^2)$ which represent the probability to find inside the hadron, in

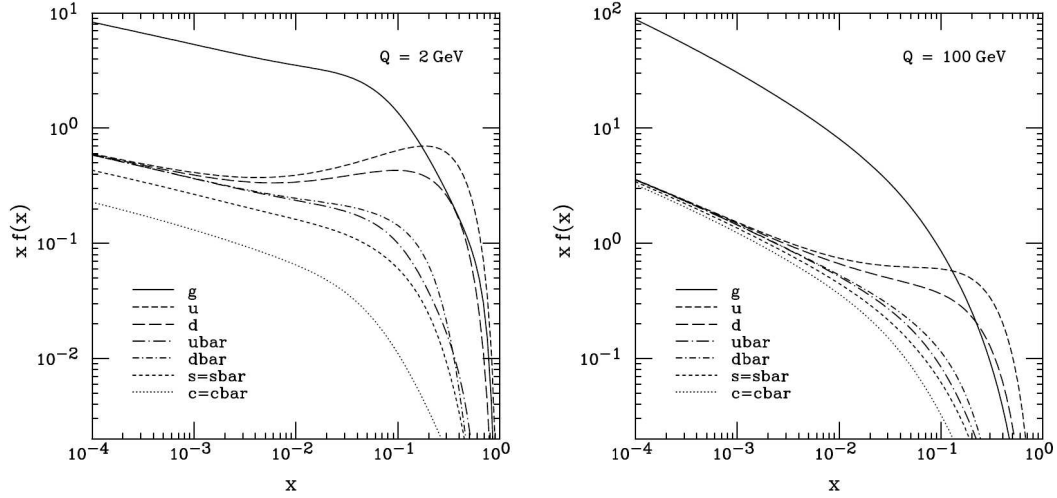


Figure 1.1: CTEQ6 parton distribution functions for $Q = 2 \text{ GeV}$ (left panel) and $Q = 100 \text{ GeV}$ (right panel).

an interaction with Q exchanged momentum, a parton of a kind, carrying a fraction of the hadron momentum in the $[x, x + \delta x]$ range. The pdf can be used to perform predictions for the interactions of the hadron with any probe. The last version of pdf provided by the CTEQ collaboration [4] is shown in figure 1.1.

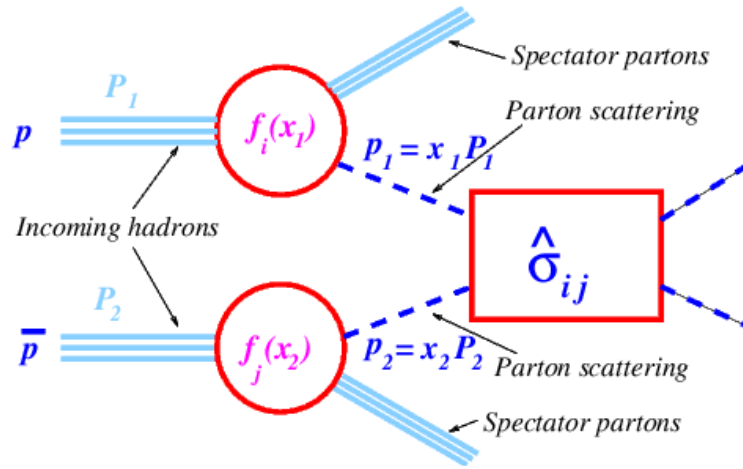
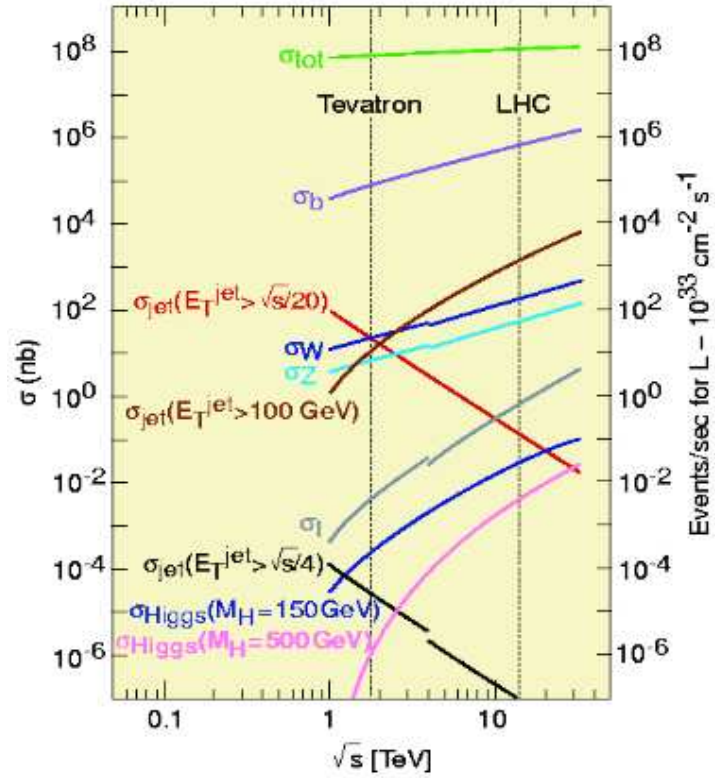
Because of the parton structure of the hadrons, a mono-energetic proton beam can be considered as a beam of partons, each of them carrying a variable fraction of the original hadron. The $p - \bar{p}$ interaction can then be interpreted as a hard interaction between two partons, as sketched in figure 1.2. The real center of mass energy ($\sqrt{\hat{s}}$) of the interaction between two partons is lower with respect to the one of the $p - \bar{p}$ collision (\sqrt{s}):

$$\sqrt{\hat{s}} = \sqrt{x_a x_b s} \quad (1.1)$$

where x_a and x_b are the proton momentum fractions carried by the two interacting partons.

The cross section of a generic process in a $p - \bar{p}$ interaction can be expressed in terms of its parton cross section $\hat{\sigma}_{ab}$:

$$\sigma = \sum_{a,b} \int dx_a dx_b f(x_a) f(x_b) \hat{\sigma}_{ab}(\sqrt{\hat{s}}) \quad (1.2)$$

Figure 1.2: Schematic representation of a parton interaction in a $p-\bar{p}$ collision.Figure 1.3: Production cross section predicted for hard scattering processes as a function of the center of mass energy for hadron colliders (pp and $p\bar{p}$).

Decay mode	Partial width [MeV]	Branching ratio[%]
e^+e^-	83.984 ± 0.086	3.3658 ± 0.0034
$\mu^+\mu^-$	84.06 ± 0.25	3.369 ± 0.010
$\tau^+\tau^-$	84.14 ± 0.28	3.372 ± 0.012
invisible	499.0 ± 1.5	19.998 ± 0.018
hadrons	1744.4 ± 2.0	69.91 ± 0.10

Table 1.1: Measured partial widths and branching ratios for the Z boson.

where the sum is over the flavours, and $f(x_q)$ represents the probability to find a q quark with a fraction x of the momentum of the parent parton.

The parton cross section for a process $q\bar{q} \rightarrow Z$ can be readily calculated in the Glashow-Weinberg-Salam model ([1], [2], [3]):

$$\hat{\sigma}_{q\bar{q} \rightarrow Z} = \frac{\pi}{3} \sqrt{2} G_F M_Z^2 (V_q^2 + A_q^2) \delta(\hat{s} - M_Z^2) \quad (1.3)$$

The partons not involved in the hard scattering interaction have small transverse momentum (P_T) and hence are distributed at small angles with respect to the beam line. In figure 1.3 the production cross section predicted for hard scattering processes in hadron collisions is reported.

1.1.1 Z boson decay

At the leading order in electroweak perturbation theory the partial widths of the Z boson in a couple of fermions ($f\bar{f}$) are given in the Standard Model by

$$\Gamma(Z \rightarrow f\bar{f}) = N \frac{G_F M_Z^3}{6\sqrt{2}\pi} (V_f^2 + A_f^2). \quad (1.4)$$

The measured values of the partial widths of the Z boson are reported in table 1.1, together with the corresponding branching ratios. Although for the Z boson the hadron decay modes are enhanced with respect to the leptonic modes, at hadron colliders there is a very serious background from normal QCD two-jet production. On the other hand, the electron and muon decay channels present a clear experimental signature and are then used to select

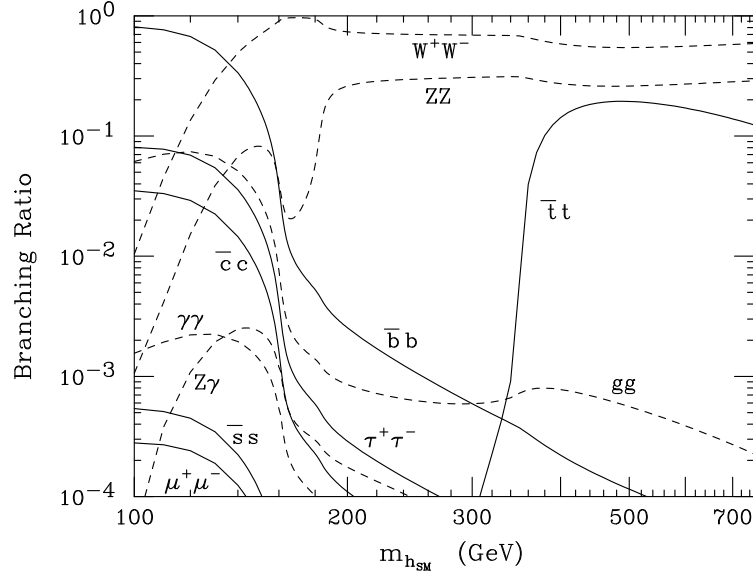


Figure 1.4: Branching ratios for the main decays of the SM Higgs boson [10].

pure samples of Z bosons with very low background. The inclusive production cross section of the Z boson, measured by the CDF collaboration in its lepton decay [5] is $\sigma_{meas}(Z, Z \rightarrow l^+l^-) = 254.9 \pm 3.3(stat.) \pm 4.6(sys.) \pm 15.2(lum.)pb$.

The prediction for inclusive Z boson production at the Tevatron energy have been computed at the NNLO and are rather precise (see for example [6], [7], [8], [9]):

$$\sigma(p\bar{p} \rightarrow Z) \cdot \text{BR}(Z \rightarrow l^+l^-) = 251.3 \pm 5.0pb$$

with uncertainty dominated by the uncertainties in the pdf set used, and agrees pretty well with the above mentioned experimental result.

1.2 Associated production of a Z boson and Heavy Flavours

Vector bosons are usually produced at colliders together with very soft QCD radiation, hence they are produced with small transverse momentum with respect to beam axis. In this region non-perturbative QCD dominates and theoretical calculations are virtually impossible. On the other hand when

W/Z recoil on high transverse momentum partons (quarks or gluons) perturbative QCD is helpful and detailed calculation up to NLO order are possible for the emission of up to two energetic partons recoiling against W or Z bosons. The experimental study of W/Z boson in association with one or more high-pt jets offers an interesting benchmark on perturbative QCD theoretical calculations. $W/Z + jets$ production is also an important background for top , Higgs and beyond-the-standard-model searches. Verifying the prediction of theoretical calculation and related Monte Carlo generators on the kinematics of this process is so important also from an experimental view point.

In particular the processes of associated production of vector bosons and heavy quarks (b or c) has received a special attention.

From an experimental point of view the $Z + b\bar{b}$ process represents the main irreducible background to the Standard Model Higgs boson produced in association with a Z boson, for low mass Higgs ($M_H < 135 \text{ GeV}/c^2$), since the $b\bar{b}$ decay channel has the highest branching ratio in this region, as shown in figure 1.4.

In addition the production of a Z boson plus two jets with at least one heavy-quark jet is also a background to the production of a Higgs boson in association with one or more b jets, which is a discovery mode for a supersymmetric Higgs boson at large values of $\tan\beta$ ¹ (see for example [11], [12]).

From a theoretical point of view this process can provide important informations on the *parton distribution functions* (*pdf*) of the Heavy Flavours inside the proton. In fact as it is apparent from the *leading order* (LO) diagrams for the associated production of a Z boson and a heavy quark (b or c), shown in figure 1.5, the $Z + b$ process can be thought as occurring via excitation of a b quark from the quark-antiquark sea inside the proton. The cross section of this process hence depends on the amount of $b\bar{b}$ in the proton *pdf*. This is a rather unconstrained part of the proton *pdf*'s and is usually derived perturbatively from the gluon *pdf* assuming a known gluon to b -quark splitting function. A

¹The minimal supersymmetric standard model requires two Higgs doublets; the ratio of their vacuum expectation values is $\tan\beta \equiv v_2/v_1$.

precise measurement of differential cross sections of $Z + hf$ processes can lead to a significantly reduced uncertainty in this *pdf*. It is worth mentioning here that the knowledge of the $b - quark$ *pdf* can significantly affect the cross section prediction for several important processes related to Higgs search at LHC, especially for SUSY at high $\tan\beta$ where coupling of Higgses to $b - quarks$ are enhanced.

Recent improvements in the calculation techniques ([13], [14] [15]) provide the *next-to-leading order* (NLO) calculation of the production cross sections for $Z + hf$ or $H + hf$ processes. By implementing the NLO calculation in the MCFM (*Monte Carlo for FeMtobarn*) simulator, Campbell, Ellis, Maltoni and Willenbrock [15] have evaluated the production cross sections for the $Z + hf$ process, according to the following prescriptions:

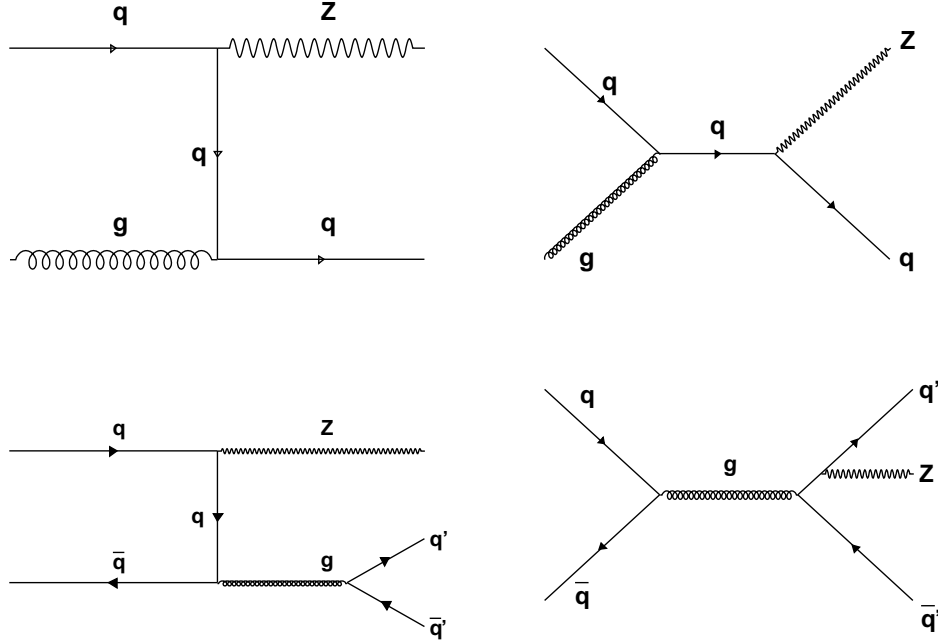
- $\sqrt{s} = 1.96$ TeV;
- jet transverse momentum $P_T > 15$ GeV/ c ;
- jet pseudorapidity $|\eta| < 2$;
- two final state partons are merged in a single jet if $\Delta R(1, 2) < 0.7$.

The obtained results are reported in table 1.2.

In the present work a study of the experimental measurement of the production cross section for events of associated production of a Z boson and at least one heavy quark is presented.

1.3 Jet physics

A peculiar feature of the strong interaction is the shape of the potential energy between quarks and antiquarks, which increases with separation: $V(r) \sim \lambda r$. This so-called “*infrared slavery*” is responsible for the total *confinement* of quarks in colorless hadrons. For this reason the partons emerging from a hard scattering interaction will generate, through a process called *hadronization*, a bunch of hadrons (mesons and/or baryons), collimated around the original parton direction. These *jets* of particles are the only measurable

Figure 1.5: *Leading order* diagrams for the $Z + hf$ production.

Cross sections (pb)		Tevatron			
	ZQ	$Z(Q\bar{Q})$	ZQj	$ZQ\bar{Q}$	ZQ inclusive
$gb \rightarrow Zb$	(8.23) 10.4	0.169	2.19	0.631	$13.4 \pm 0.9 \pm 0.8 \pm 0.8$
$q\bar{q} \rightarrow Zb\bar{b}$	3.32	1.92		1.59	6.83
$gc \rightarrow Zc$	(11.3) 16.5	0.130	3.22	0.49	$20.3^{+1.8}_{-1.5} \pm 0.1^{+1.3}_{-1.2}$
$q\bar{q} \rightarrow Zc\bar{c}$	5.66	6.45		1.70	13.8
	Zj		$Zj\bar{j}$		Zj inclusive
$q\bar{q} \rightarrow Zg, gq \rightarrow Zq$	(876) 870		137		$1010^{+44}_{-40} {}^{+9}_{-2} {}^{+7}_{-12}$

Table 1.2: Production cross section of the associated production of a Z boson and hadronic jets, divided according to the flavour of the parton produced and to the experimental signature: $ZQ = Z$ and exactly one heavy-quark jet; $Z(Q\bar{Q}) = Z$ and exactly one jet containing two heavy quark; $ZQj = Z$ and two jet, only one containing a heavy quark; $ZQ\bar{Q} = Z$ and two jets, both containing a heavy quark; $Zj = Z$ and a jet not containing a heavy quark; $Zj\bar{j} = Z$ and two jets both not containing a heavy quark. In parenthesis the *leading order* is reported.

objects which provide information on the parent parton originating from the hard scattering.

In order to extract information on the dynamics of the event and make comparison between experimental data and theoretical predictions, a quantitative jet reconstruction algorithm has to be defined [16].

A jet algorithm should satisfy some basic requirements. It should be simple to be applied in the experimental framework and it has to be infrared safe. The latter property is to regularize the divergent contributions which arise in the perturbative calculation of cross sections. In this attempt the divergent soft and collinear radiation, whose contribution cancel out with the virtual effects, has to be treated as not resolvable, and its energy summed with the parent parton energy. From the experimental point of view, this minimal parton resolution is represented by the calorimeter cell segmentation; two particles hitting the same cell are treated as if they were a single object. Jet algorithms should also be stable, and able to treat soft energy radiation generated during the parton shower, without affecting the macroscopic properties of jets.

In the present analysis the jets are reconstructed using the *cone algorithm*, which is briefly described in section 1.3.1.

1.3.1 The cone algorithm

The cone algorithm relies on the assumption that particles close in space are originated from the same parton [16]. Particles which lie in a circle of given radius R in $\eta - \phi$ space are merged together. The procedure starts constructing a cluster of particles around a trial centroid (*seed*); subsequently the energy-weighted center of the jet cone is calculated, and a new cluster is built around the new centroid. This procedure is iterated until the stability of the jet configuration is reached. The particles are specified by massless

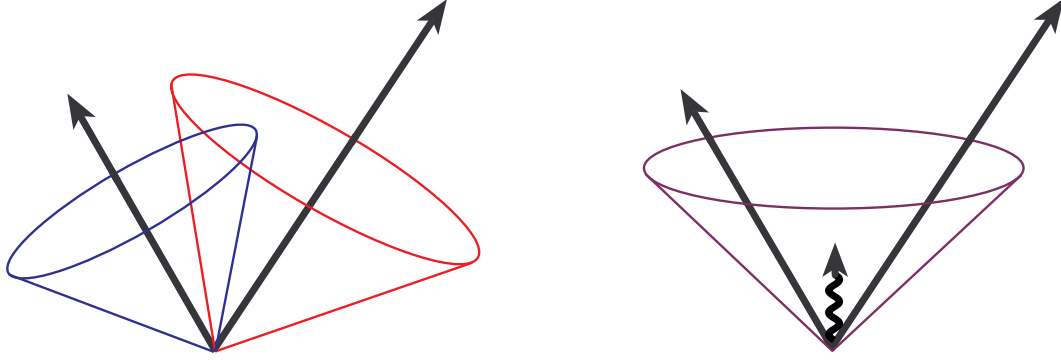


Figure 1.6: An illustration of infrared sensitivity in cone jet clustering. In this example, jet clustering begins around seed particles (left panel), shown here as arrows with length proportional to energy. The presence of soft radiation between two jets (right panel) may cause a merging of the jets that would not occur in the absence of the soft radiation.

four-vectors with directions (η, ϕ) , the jet four-vector is then defined by:

$$E_T^{jet} = \sum_p E_T^p \quad (1.5)$$

$$\eta^{jet} = \frac{1}{E_T^{jet}} \sum_p E_T^p \eta^p \quad (1.6)$$

$$\phi^{jet} = \frac{1}{E_T^{jet}} \sum_p E_T^p \phi^p \quad (1.7)$$

A schematic example of the clustering evolution according to the cone algorithm is reported in figure 1.6.

1.4 Monte Carlo generators

The $Z + jet$ cross section calculation involves many subprocess contributions and depends on the jet definition. Beside the complexity of the parton level cross section, the final result is affected by the algorithm used to develop the parton shower, the hadronization and the jet clustering. Several Matrix Element (ME) programs are today available to generate $p\bar{p} \rightarrow Z + jets$ events. These generators, based on a full ME calculation at some given perturbation

order in the coupling constants, properly describe the hard, large-angle emission of partons. This capability is crucial to have a precise kinematic picture of the process under study. To have a full event description in terms of physical particles, one should evolve the generated partons up to the fragmentation scale at which the hadronization process takes place. This stage of the evolution of an event is performed by the Parton Shower (PS) programs, that, in the so-called soft and collinear approximation, describe the physical evolution of a parton between the hard energy scale (the one used in the ME calculation) and the fragmentation scale. During this evolution the hard parton loses energy by radiating quarks and gluons until fragmentation takes over. After hadronization (and decay of unstable particles) the full event picture is represented by jets of particles close to the initial parton directions. These particles are then merged into physical jets using a jet clustering algorithm. To address a complete and efficient Monte Carlo event generation several effects have to be taken into account. Beside the parameter tuning and the choice of inputs, there are conceptual problems arising from the combination of ME and PS programs. Since both the ME and PS approaches deal with parton radiation, in some circumstances some final topologies, defined by the jet multiplicity, may receive contribution from the hard process as well as from the shower evolution which gives rise to double counting problems.

In fact, after the parton shower, the fragmentation process takes over and the final state will appear as bunches of hadrons near the original parton directions. The number of jets generated following these steps is however not in one to one correspondence with the number of partons produced with the ME calculation. The parton shower may change the jet multiplicity generating new jets from initial and final state radiation, or smear the initial parton energy such that it is not any more reconstructed as a jet.

The way in which all the ME programs are interfaced with all the PS programs in hadronic interaction at the moment does not give a definite prescription about how to divide the available phase space into two regions, one covered by the ME calculation and the other by the PS algorithm, and so

double counting problems might arise as a natural consequence.

In fact, even if some parton configurations might not pass the reconstruction cuts at particle level, they could still give rise to hadron final states surviving after final jet cuts. The shower evolution, the presence of underlying events and the detector simulation could in fact slightly change the kinematic of the event. As an example can be considered a configuration in which two partons are close enough to be reconstructed as a unique jet, and a gluon radiated by the PS is hard enough to be reconstructed as a jet. This topology is exactly what is to be avoided: the hard large-angle emission should be covered by the ME by evaluating properly all the interferences and the angular correlations. Naively speaking, one might think that these topologies occur with a low frequency roughly proportional to an extra power of α_S due to the additional gluon leg. Nevertheless there are configurations, like the collinear one, in which the ME is sensitive to some divergences due to the leading order nature of this calculation. The occurrence of this configurations is then enhanced by some large unphysical weight, that combined with the low probability to emit an extra parton, may give rise to a non-negligible contribution distorting the total cross section as well as the distributions. This effect is clearly visible in the sensitivity of the final cross section to the parton level cuts: the closer one is to the divergent regions, the bigger is the jet cross section.

1.4.1 Alpgen

The ALPGEN Monte Carlo program [17] bases the ME calculation on the Alpha algorithm [18]. The Alpha code returns to the main program the value of the ME including all the effects coming from spin and color correlation as well as mass effects. Due to the complexity of the multi-parton final state processes the integration over the phase space is done numerically in a Monte Carlo fashion. The problem that the behavior of the cross section on the available phase space is not known *a priori*, results in a quite large event generation inefficiency which limits the speed of the calculation.

The calculation of the cross section proceeds through the following steps:

- The hard process is set up, defining the parton multiplicity and the kinematic limits on the phase space.
- A first set of phase space integration cycles is performed in an attempt to parametrize how the cross section is distributed in phase space and among the possible subprocesses. On an event by event basis, the following steps take place:
 - a subprocess is randomly selected;
 - according to the kinematic cuts a point in the phase space is randomly selected;
 - the initial state parton luminosity is evaluated and a flavour configuration is selected;
 - spin and color for each parton are randomly assigned;
 - the matrix element is evaluated, which, including the phase space and the luminosity factor, determines the weight of the event.
- This procedure is iterated for a few cycles until a stable map of the cross section distribution over the phase space is obtained. At each step the event generation is weighted using the map calculated in the previous iteration.
- After the selected number of iteration is completed a final long-statistic run is performed using the optimized phase space grid obtained in the warm-up cycles.

The cross section is obtained by summing over all the weights generated. Moreover, thanks to the Monte Carlo nature of the calculation, the output also contain the list of event generated with the respective kinematic characterization. In order to have a sample of events with the same relative weight suitable for further studies, events can be unweighted at the end of the generation. The unweighting procedure simply means that events are kept or rejected according to their probability to be generated. This probability is defined by the ratio of the weight of each event and the maximum weight found in the event sample.

In order to allow parton shower evolution of these events, each event is assigned a particular color configuration [19].

After the ME unwheigting, events are processed by parton shower Monte Carlo (like Pythia [20] or Herwig [21]), evolving partons from the hard-scattering scale to the fragmentation scale in the parton shower approach.

The Alpgen ME calculation is combined with the parton shower according to the *MLM matching* prescription: partons are clustered after the shower and compared to the partons produced by the ME; events in which a parton produced by ME calculation does not match a unique jet are rejected.

Chapter 2

The Tevatron and the CDF Detector

High Energy Physics (HEP) can be studied in cosmic ray interactions, where the primary particle energy can reach 10^{20} eV . However, events at the highest energies are extremely rare and can not be controlled by the scientists. High energy interactions of nuclear particles can be produced in the laboratory by means of particle accelerators, providing a much higher event rate and under much cleaner experimental conditions, albeit at less extreme reaction energies.

The largest particle accelerator ever built was LEP, an e^+e^- ring which reached the C.M.S¹ energy $\sqrt{s} = 205 \text{ GeV}$ running in a 27 km circular underground tunnel close to the city of Geneva (Switzerland). After the shut-down of LEP at the end of the year 2000, the largest machine presently in operation is the **Tevatron**, a $p\bar{p}$ collider reaching the energy of $\sqrt{s} = 1.96 \text{ TeV}$ in a 6.4 Km ring. The Tevatron is located about 50 km west of Chicago (USA) in Fermilab, a scientific laboratory run by a consortium of universities (URA) and by the University of Chicago ("Fermi Research Alliance") on behalf of the American Department of Energy (DOE).

Two detectors along the Tevatron collider collect physics events: CDF and DØ. After the first evidence for the existence of Top quark shown by CDF in 1994, the CDF and DØ Collaborations announced the discovery the Top

¹Center of Mass System.



Figure 2.1: An airplane view of the Fermilab laboratory. The ring at the bottom of the figure is the Main Injector, the above ring is the Tevatron. On the left are clearly visible the paths of the external beamlines: the central beamline is for neutral beams and the side beamlines are for charged beams (protons on the right, mesons on the left).

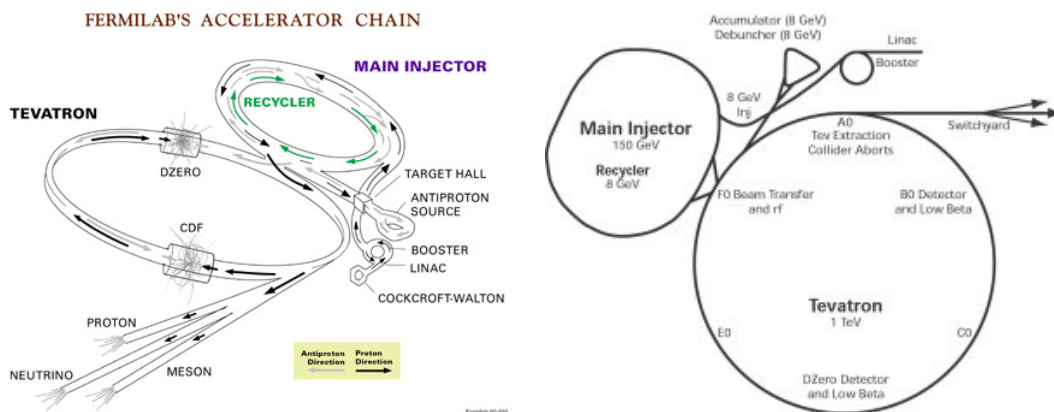


Figure 2.2: Layouts of the acceleration chain at Fermilab Tevatron, from the source to the collision.

Run	Period	Int Lum pb^{-1}
First Test	1987	0.025
Run 0	1988-1989	4.5
Run 1A	1992-1993	19
Run 1B	1994-1995	90
Run 1C	1995-1996	1.9
Run 2	2001-	>4000

Table 2.1: Integrated luminosity delivered by Tevatron in its physics runs. Run 2 is still in progress.

quark in 1995.

At present (year 2008) the Large Hadron Collider (LHC) is being built at CERN in the LEP tunnel. This new machine will start operating in late 2008 and will eventually reach a C.M.S energy of $\sqrt{s} = 14 \text{ TeV}$.

2.1 The Tevatron

The Tevatron is the final and largest element of the Fermilab accelerator complex, whose structure is illustrated in figure 2.2. The Tevatron works primarily as a $p\bar{p}$ collider. However, it can also accelerate a single proton beam and operate in fixed target mode to provide a number of neutral and charged particle beams. The Tevatron collider obtained the first collisions in 1985. In the course of time it provided several physics runs as listed in table 2.1.

2.1.1 The proton beam

The particle beams circulating in the Tevatron are provided by a chain of accelerators that use different techniques to extract, collect and accelerate protons.

- **H^- source:** a 25 KeV H^- electrostatic source provides negative hydrogen ions.

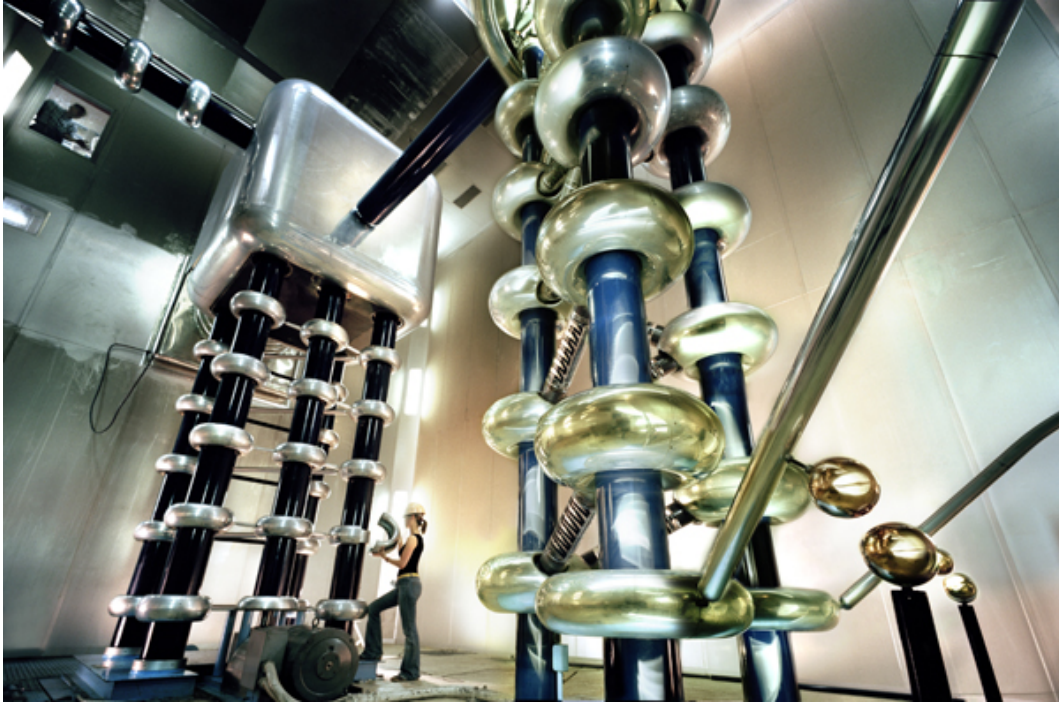


Figure 2.3: The Cockcroft-Walton accelerator is the starting point of the proton acceleration chain.

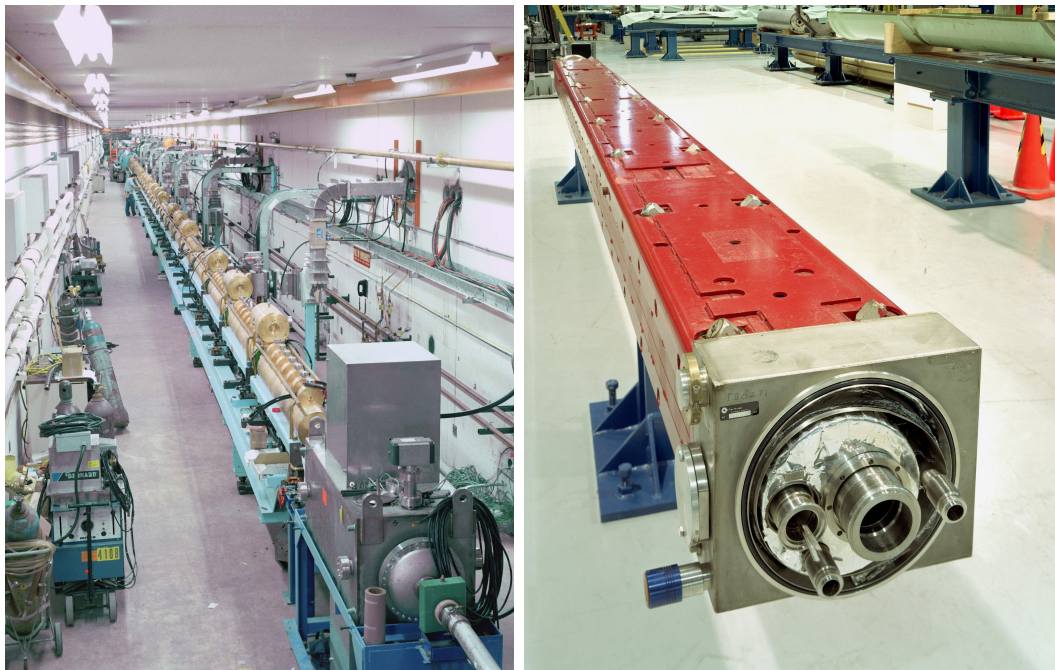


Figure 2.4: Left: upstream view of the 400 MeV section of the Linac. Right: Tevatron Superconducting Dipole Magnet.

- **Cockcroft-Walton electrostatic accelerator:** the particle acceleration cycle begins with a Cockcroft-Walton electrostatic accelerator (see figure 2.3) feeding a linear accelerator with H^- at an energy of 750 *KeV*.
- **Linac:** the Linac accelerates H^- ions up to 400 *MeV* energy. The linac was upgraded to this energy in 1993 when the final energy was doubled, as well as the number of protons per bunch. Figure 2.4 (left) shows a portion of the Linac accelerator.
- **Booster:** H^- ions are stripped at the linac exit through a carbon foil and bare protons are delivered to the Booster. This is a 8 *GeV* synchrotron with a 150 *m* diameter. The Booster transfers protons to the Main Injector with an efficiency of $\sim 75\%$. Figure 2.5 shows the Booster tunnel.

2.1.2 Main Injector

The Main Injector, completed in 1999 for Run II, is located in a 3 *Km* circumference tunnel (which houses also the antiproton Recycler, see section 2.1.3 and figure 2.6), and is approximately tangent to the Tevatron. The Main Injector functions can be summarized as follows.

- It accelerates protons from 8 *GeV* to 150 *GeV* and feeds them to the Tevatron (in the clockwise direction).
- In the antiproton production phase it accelerates protons to 120 *GeV*. Protons are extracted and directed to the antiproton production target. Antiprotons are accumulated and cooled at 8 *GeV* in the Antiproton Source.
- It receives antiprotons from the Antiproton source, boosts their energy to 150 *GeV* and transfers them to the Tevatron (in the anti-clockwise direction).

The Main Injector maximum stored beams are $\sim 3 \cdot 10^{13}$ protons and $\sim 2 \cdot 10^{12}$ antiprotons. Beams are stored in 36 bunches in the Tevatron.



Figure 2.5: The Booster tunnel. A klystron is visible in the center of the picture.



Figure 2.6: Main Injector (blue magnets on bottom) and the Recycler (green magnets on top) in the Main Injector tunnel.

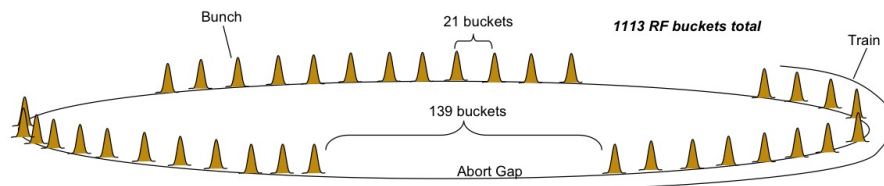


Figure 2.7: Bunch structure of the Tevatron beams in Run II.

2.1.3 Antiproton Production and Recycler

Antiprotons (\bar{p}) are produced from the 120 *GeV* proton beam extracted from the Main Injector and focused on a nickel target.

Antiprotons are collected at 8 *GeV* with wide acceptance around the forward direction, injected into a Debuncher Ring, debunched into a continuous beam and stochastically cooled. The beam is then transferred between cycles (about 2" long) to the Accumulator where antiprotons are stored at a rate of about $25 \cdot 10^{10} \bar{p}/hour$ (improvements in the storage rate are still being made).

Stacking within the accumulator acceptance is limited to a stored beam of about 10^{12} antiprotons. The accumulated beam is then transferred to the large acceptance Recycler, an 8 *GeV* permanent magnet ring housed in the Main Injector tunnel which has an acceptance about twice as large as the accumulator. In normal conditions every 30' the Recycler receives about $2 \cdot 10^{11}$ antiprotons from the Accumulator with a current of about 10 *mA*. Recently (2005) the *electron-cooling* technique was successfully applied to the Recycler to cool the antiproton beam by a 4.3 *MeV* electron beam, which is provided by a Pelletron accelerator adjacent to the ring.

2.1.4 Tevatron

The Tevatron is a 1 *Km*-radius circular synchrotron employing superconducting bending magnets (figure 2.4, right), where the protons and antiprotons beams orbit in the same pipe in opposite directions. Undesired bunch crossings are avoided by electrostatic separators.

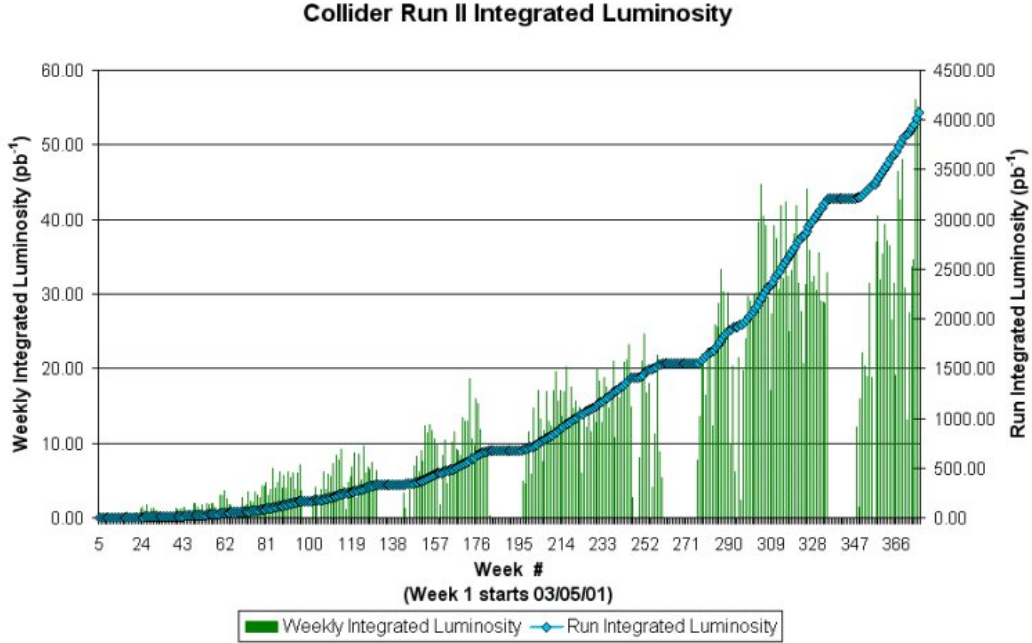


Figure 2.8: The integrated luminosity in Run II weeks. The empty periods of time correspond to Tevatron shutdowns.

The Tevatron receives protons and antiprotons at 150 GeV and ramps² them to 980 GeV (in Run II) where they are kept circulating in opposite directions at constant energy for physics runs lasting up to ~ 30 hours. Stable running conditions and data-taking by the experiments are reached after beams are scraped with remotely-operated collimators to remove the beam halo.

The beam revolution time is 21 μs . The beams are split in 36 bunches organized in 3 trains each containing 12 bunches (see figure 2.7). Within a train the time spacing between bunches is 396 ns . An empty sector 139 buckets-long (2.6 μs) is provided in order to allow the kickers to raise to full power and abort the full beam into a dump in a single turn. This is done at the end of a run or in case of an emergency.

²The magnetic field is ramped up together with the energy in order to maintain the revolution radius constant. The final condition when the magnetic field is maximum and is kept constant is called *flat-top*.

During data-taking runs the luminosity decreases approximately exponentially with time. The record initial luminosity reached by the Tevatron as of March 2008 has been about $315 \cdot 10^{30} \text{ cm}^{-2} \text{ s}^{-1}$. Figure 2.8 shows the weekly integrated luminosity during the Run II period as a function of time.

At the end of a run, when the luminosity is too low to allow a significant data-taking (typically $20 \cdot 10^{30} \text{ cm}^{-2} \text{ s}^{-1}$), the beams are aborted and the *shot-setup* procedure is started to prepare for a new store. In optimal conditions runs last about 30 hours.

2.2 The CDF Detector in Run II

The CDF detector described here below is as configured for Run II. A schematic view of the detector is presented in figures 2.9 and 2.10.

The CDF architecture is quite common for this type of detectors. Radially from the inside to the outside it features a tracking system contained in a superconducting solenoid, calorimeters (electromagnetic and hadronic) and muon detectors. The whole CDF detector weights about 6000 tons.

2.2.1 System Reference

The CDF detector is approximately cylindrically symmetric around the beam axis. Its geometry can be described in cartesian as well as in cylindrical coordinates.

The left-handed cartesian system is centered on the nominal interaction point with the \hat{z} axis laying along the proton beam and the \hat{x} axis on the Tevatron plane pointing radially outside.

The cylindrical coordinates are the azimuthal angle φ ($\varphi = 0$ on the \hat{x} direction) and the polar angle ϑ ($\vartheta = 0$ along the positive \hat{z} axis):

$$\varphi = \tan^{-1} \frac{y}{x} \quad \vartheta = \tan^{-1} \frac{\sqrt{x^2 + y^2}}{z} \quad (2.1)$$

A momentum-dependent particle coordinate named *rapidity* is also commonly used. The rapidity is defined as:

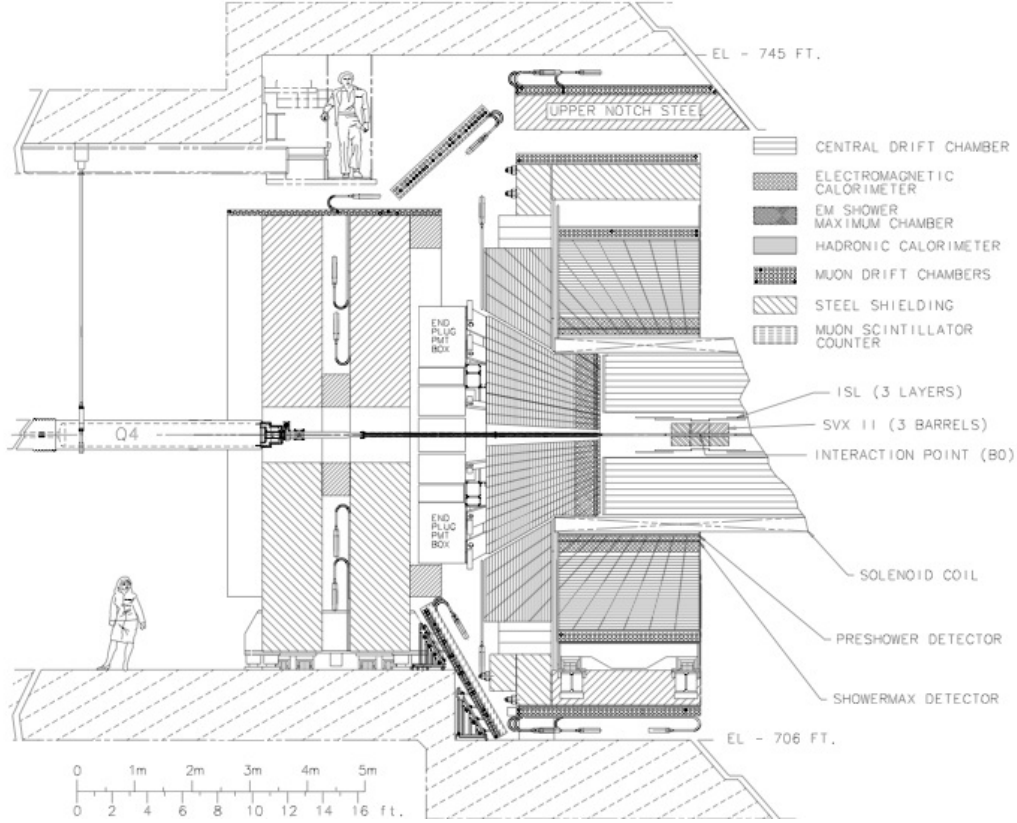


Figure 2.9: Elevation view of the CDF II detector.

$$Y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \quad (2.2)$$

Relative particle distances in rapidity are invariant under Lorentz boosts along the z axis.

The *pseudorapidity* η , is the limit of the rapidity in case of massless particles, and it depends only on the polar angle:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) \quad (2.3)$$

The pseudorapidity is commonly used to identify different detector regions according to their position respect to the beamline and interaction vertex position, as shown in figure 2.9.

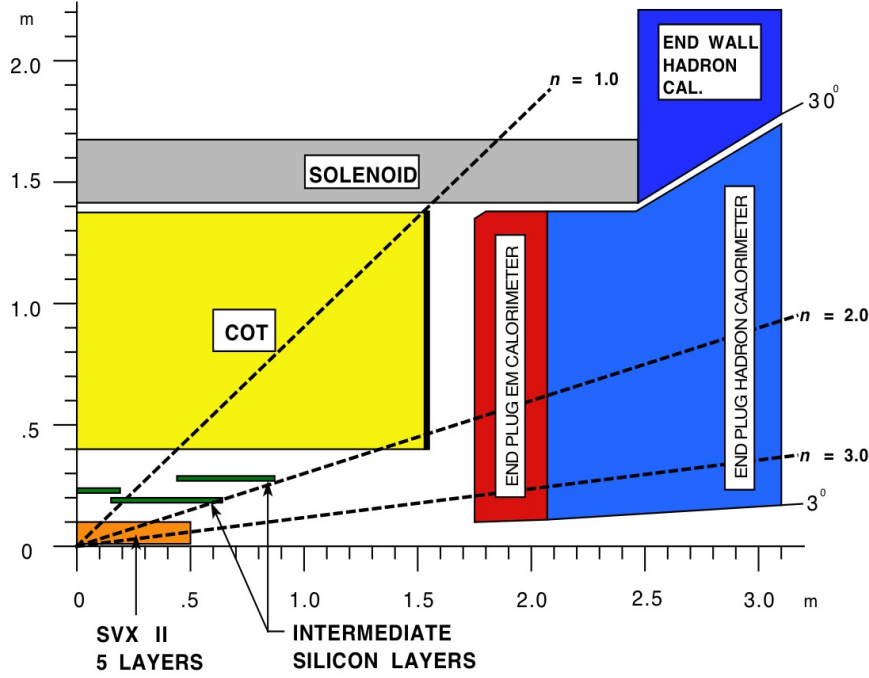


Figure 2.10: The CDF II detector projected on the z/y plane.

2.2.2 The Tracking System

The CDF tracking system is immersed in the 1.4 T magnetic field of a solenoid parallel to the beams. Particles are bent depending on their charge sign and transverse momentum.

There are two tracking systems at CDF, an inner Silicon tracker and a Central Outer Tracker (COT) drift chamber (figures 2.9 and 2.10).

Silicon Tracker

The silicon system (figure 2.11) employs stripped silicon wafers of an excellent space resolution of $\sim 12\text{ }\mu\text{m}$. It was designed to sustain the large radiation dose to be integrated in a long running period. Simulations predict that the performances of the silicon tracker will remain unaffected up to $\sim 5\text{ fb}^{-1}$ integrated luminosity.

The full CDF Silicon Detector is composed of three approximately cylindrical coaxial subsystems: radially towards the outside, the $L00$ (Layer 00), the

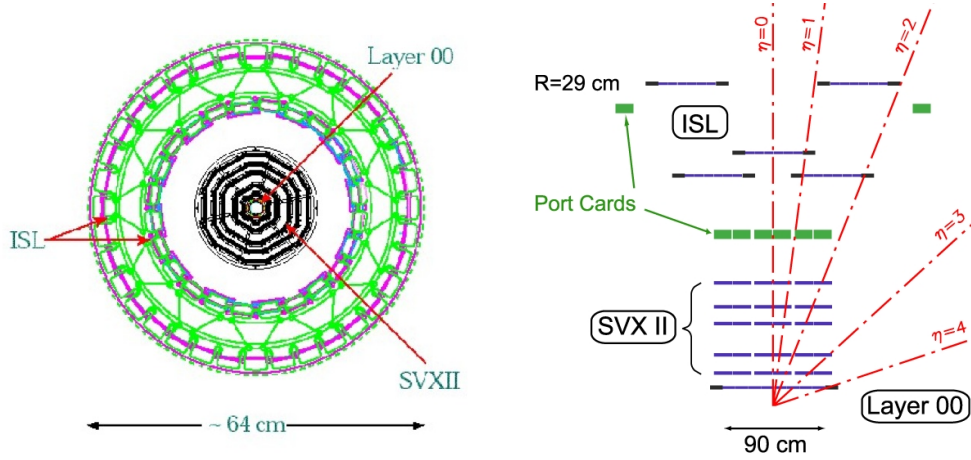


Figure 2.11: Left: cutaway transverse to the beam of the three subsystems of the silicon vertex system. Right: sketch of the silicon detector in a z/y projection showing the η coverage of each layer.

SVX (Silicon VerteX), the ISL (Intermediate Silicon Layer). A total of 400k signal channels are provided from the silicon detector.

L00 is a 90 cm long, radiation hard, single sided silicon detector, structured in longitudinal strips. It is mounted directly on the beam pipe at 1.35 – 1.62 cm from the beam axis. The detector support structure is in carbon fiber with integrated cooling system. The sensors are silicon wafers 250 μm thick with inprinted strips with 0.25 μm technology.

Being so close to the beam, L00 allows to reach a resolution of $\sim 25/30$ μm on the impact parameter of tracks of moderate p_T , providing a powerful help to identify long-lived hadrons containing a b quark.

L00 is backed by the **SVX**, a set of three cylindrical barrels 29 cm long each along z (see figure 2.12). Barrels are radially organized in five layers of double-sided silicon wafers extending from 2.5 cm to 10.7 cm. Three of those layers provide φ measurement on one side and $90^\circ z$ on the other, while the other two provide φ measurement in one side and a z measurement by small angle 1.2° stereo on the other.

The SVX detector has $\sim 90\text{cm}$ of total active length, which corresponds to about 3σ of the gaussian longitudinal spread of the interaction point, and

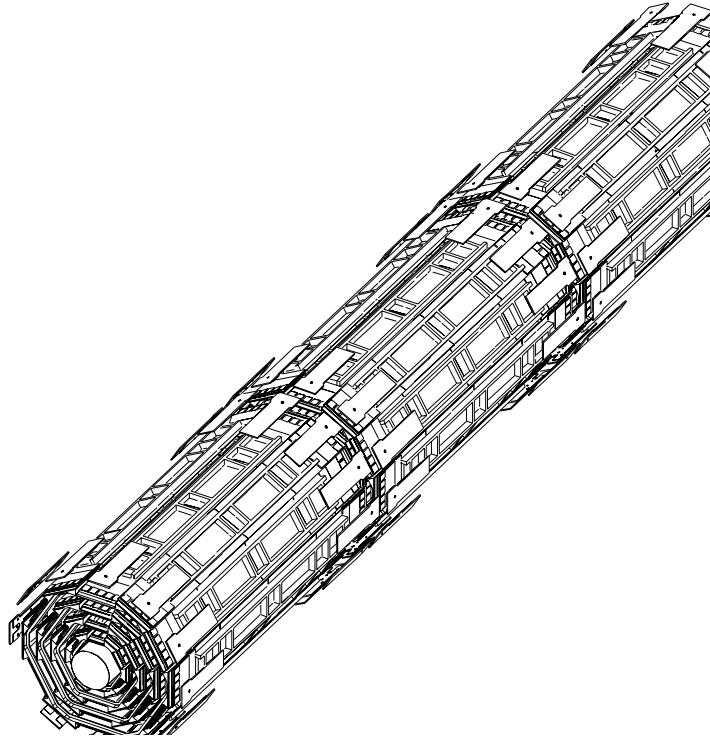


Figure 2.12: The SVX silicon detector consists of three barrels with five layers of silicon detectors each.

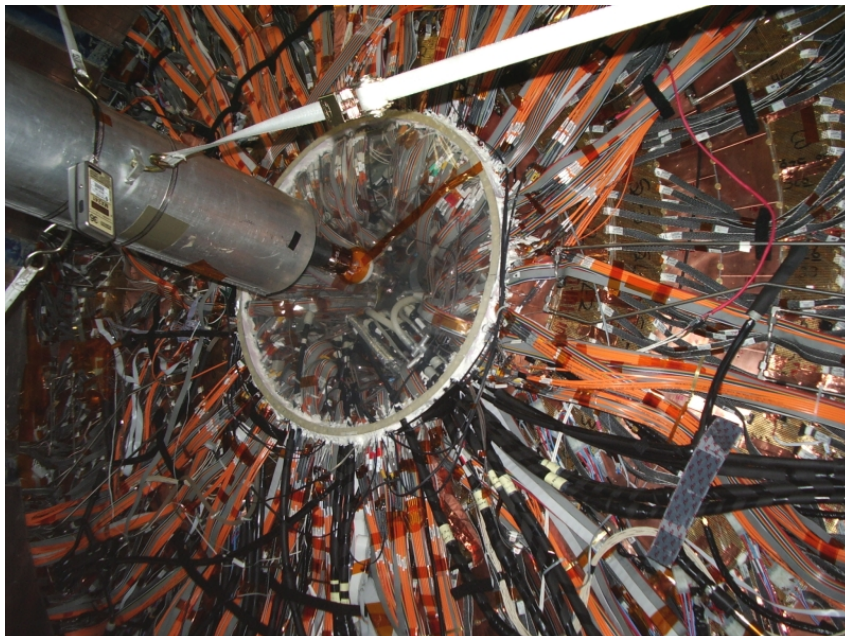


Figure 2.13: The beam pipe at SVX entrance. Well visible the cooling pipes and cables carrying power and detector signals.

provides pseudorapidity coverage in the $|\eta| < 2$ region.

The SVX detector plays a crucial role both for the offline reconstruction and for the B hadrons trigger (*SVT*), which is based on the online identification of the secondary vertexes displaced by the primary interaction point.

The **ISL** consists of 5 layers of double sided silicon wafers: four are assembled in two telescopes at 22 *cm* and 29 *cm* radial distance from the beamline covering $1 < |\eta| < 2$. One is central at $r = 22$ *cm*, covering $|\eta| < 1$. The two ISL layers at $1 < |\eta| < 2$ are important to help tracking in a region where the COT coverage is incomplete.

COT

The Central Outer Tracker (COT) is a 310 *cm* long open-cell drift chamber, positioned at $43 < r < 137$ *cm* radial distance just outside the ISL and covering the pseudorapidity range $|\eta| < 1$. The COT detector provides precision tracking and transverse momentum (p_T) measurement for charged particles.

It has 4 axial and 4 stereo³ *superlayers*, alternating with each other and consisting of 12 wire layers each for a total of 96 layers and over 30k wires. The sense wires of each layer are alternated with field shaping wires. The chamber is filled with a fast gas mixture (Ar-Ethane-*CF*₄).

In the COT solenoid magnetic field the electrons drift at $\simeq 35^\circ$ (Lorentz angle) with respect to the direction of the cell electric field. The resolution in the transverse to radial direction is maximized by tilting the cell by the same angle to make the drift path perpendicular to the radius. This geometry also makes high p_T track crossing cells at different relative distances from sense wires thereby reducing the systematic errors due to drift field distortions within a cell. The single hit position resolution has been measured to be ~ 140 μm which translates into an overall transverse momentum resolution of:

$$\frac{\partial p_T}{p_T} = 0.15\% \frac{p_t}{[GeV/c]} \quad (2.4)$$

³The stereo wires are tilted at $\pm 2^\circ$ with respect to the z direction.

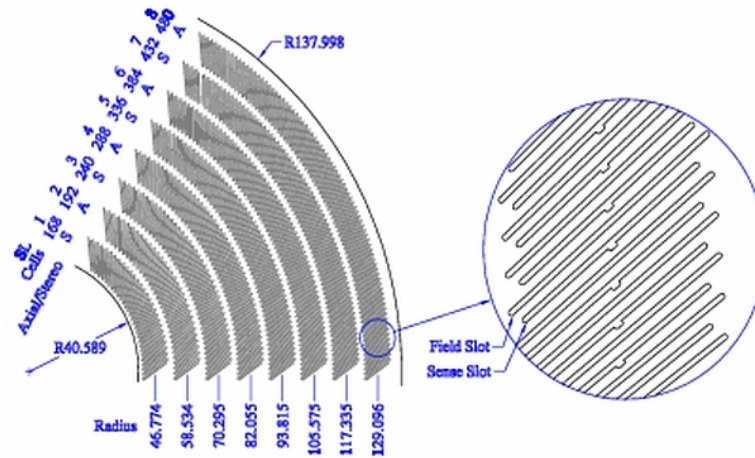


Figure 2.14: One sixth of the COT endplate. The wires are grouped in 8 superlayers.

2.2.3 Time of Flight

The Time of Flight (TOF) detector is a recent upgrade of the CDF detector. TOF is an array of approximately 4 *cm* thick and wide, 279 *cm* long scintillator bars⁴ parallel to the beam, which is laied as a cylindrical sheet of 216 elements on the inner solenoid wall, at a radial distance $r = 138$ *cm*. Photomultiplier tubes, that are capable to provide adequate gain even in presence of the magnetic field, are connected at both ends of the bars and provide time and pulse hight measurements. By comparing the two measurements, the z coordinate can also be determined.

A resolution of ~ 110 *ps* has been achieved which allows a 2σ separation of kaons from pions up to ~ 1.6 *GeV* at $|\eta| < 1$.

2.2.4 The Solenoid

The CDF detector is built around a solenoid providing a longitudinal magnetic field of 1.4 *T*. All the tracking system is inside this field. The solenoid volume is a cylinder 3.5 *m* long and 2.8 *m* in diameter (see figure 2.10).

⁴Bars have a slightly trapezoidal-cross section. They are so shaped in order to minimize particle losses through the cracks between bars.

The coil is an Al-stabilized NbTi superconductor operating at liquid helium temperature. The operation current is 4660 A. After careful cool-down procedures during Run 1, the solenoid could be reused in Run II.

The solenoid radial thickness corresponds to $1.075 X_0$ and initiates EM showers of crossing electrons and photons that are sampled by a pre-radiator detector (the CPR2 system, see section 2.2.5).

2.2.5 CPR2: the Central Preshower system

The solenoid coil is $1.075 X_0$ and as such initiates the particle showering to be sampled and integrated fully in the outer calorimeters. Radially behind the coil, in front of the calorimeters, a scintillator layer acts as a Central Pre-Radiation detector (CPR) [22] for electrons and photons. A Central Crack Radiation detector (CCR) extends the preshower to the mechanically intrigued regions between calorimeter wedges. The 2 cm thick CPR scintillators provide a clear signature of the electromagnetic showers initiated in the solenoid coil. Information from the CPR is also useful to complement the calorimeter response for better jet energy resolution.

Figure 2.15 (left) shows one CPR2 module just before the top cover was installed. The figure shows the individually wrapped tiles and fibers exiting them.

2.2.6 Calorimeters

The CDF calorimetric system is designed to absorb up to $\sim 98\%$ the hadronic and electromagnetic energy over most of the solid angle⁵. The calorimeters are split into cells projecting from the nominal interaction point in order to associate the single cell response to energy flow in a solid angle bin.

The CDF calorimeters are sandwiches of active and converter material covering two large η regions: the Central Calorimeter covers $|\eta| < 1.1$ and the Run II plug calorimeters cover $1.1 < |\eta| < 3.64$.

⁵A really 4π coverage is not possible. The beam pipe aperture cannot be covered and additional dead regions cannot be avoided at the edges of calorimeter wedges and cells.

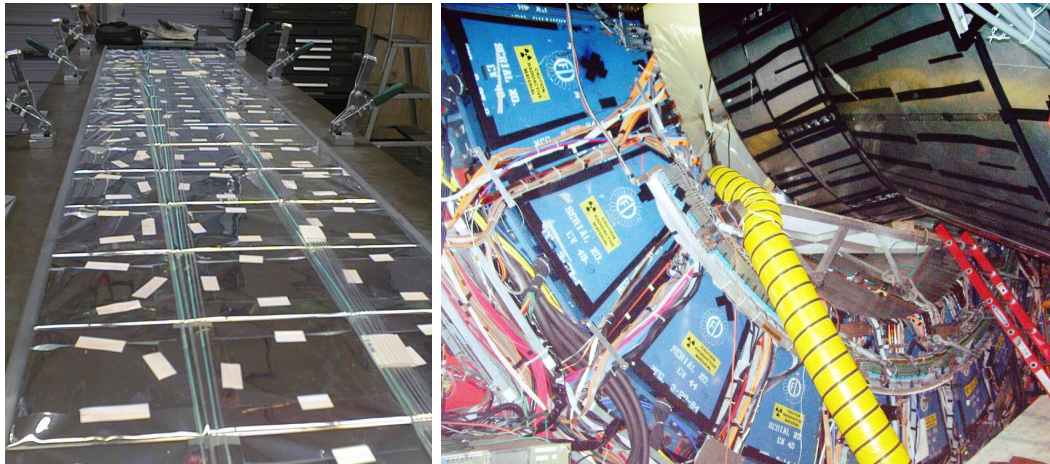


Figure 2.15: Left: one CPR module during assembly in summer 2004. Tiles are wrapped in reflecting paper to avoid cross-talks and to maximize the light collection. The clear fibers are merged together into four groups at the end of the module and connected to R5900 photomultipliers. Right: the Plug calorimeter being removed away from the main detector body. A number of elements of the wall hadron calorimeter (in blue) are also visible.

Particles coming from the interaction point cross only relatively light detectors (the tracking system and the solenoid coil) before entering the calorimeters where they start showering. The scintillator light signals are collected separately from the front lead-scintillator electromagnetic and from the rear iron-scintillator hadronic compartments. Both signals are roughly proportional to the released energy in the shower. Accurate calibrations of response to particles of known energy and detailed MC simulations are necessary to reconstruct the incoming particle and jet energy.

Central

The Central Calorimeter has been preserved intact from Run I to Run II. It consists of two coaxial barrels (east and west), each divided in two arches (left and right around the 90° polar angle). Each arch is split into 12 azimuthal wedges of 15° aperture. Finally, each wedge is split into 10 projective towers of width $\delta\eta \simeq 0.11$.



Figure 2.16: A CEM/CHA wedge during assembly. The absorber black iron frames are clearly visible as well as the light guides transporting the light from the scintillators to the photomultipliers (housed in tubes on the top of figure).

All calorimeters are radially split into an electromagnetic compartment, with lead as converter, and an hadronic compartment, with iron as converter. The light emitted in the plastic scintillator is collected in wavelenght shifting bars (central calorimeter) or fibers (plug calorimeter) and transported to the PMT's.

The CEM electromagnetic compartment has 31 layers of 5.0 *mm* polystyrene scintillator alternate with 2.5 *mm* thick lead tiles. The Wall Hadron calorimeter (see figure 2.15, right) comprises 24 elements which are embedded in the front walls of the detector body. At variance with the central calorimeter these detectors are part of the magnetic field return yoke.

The Central and Wall Hadronic Calorimeters (CHA, WHA) use iron as radiator. The CHA (WHA) has 32 (15) layers, and each layer is composed of 2.5 (5.1) *cm* of iron absorber and 1.0 (1.0) *cm* of plastic scintillator. The total calorimeter thickness is $\sim 4.7\lambda_0$ (λ_0 is the absorption length) for both CHA

	CEM	CHA	WHA	PEM	PHA
η coverage	< 1.1	< 0.9	$0.7 < \eta < 1.3$	$1.3 < \eta < 3.6$	$1.3 < \eta < 3.6$
n. of modules	48	48	48	24	24
η towers/mod	10	8	6	12	10
n. of channels	956	768	676	960	864
Absorber (mm)	Pb (3.0)	Fe (25.4)	Fe (50.8)	Pb(4.6)	Fe (50.8)
Thickness	$19X_0, 1\Lambda_0$	$4.5\Lambda_0$	$4.5\Lambda_0$	$21X_0, 1\Lambda_0$	$7\Lambda_0$
Position res.	0.2×0.2	10×5	10×5		
Energy res.	$\frac{13.5\%}{\sqrt{E_T}} \oplus 1.7\%$	$\frac{75\%}{\sqrt{E_T}} \oplus 3\%$	$\frac{80\%}{\sqrt{E_T}}$	$\frac{16\%}{\sqrt{E_T}} \oplus 1\%$	$\frac{80\%}{\sqrt{E_T}} \oplus 5\%$

Table 2.2: Geometry, parameters and performance summary of the CDF Calorimetric System. The position resolution is given in $r \cdot \phi \times z \text{ cm}^2$ and is measured for a 50 GeV incident particle.

and WHA.

Figure 2.16 shows one CEM/CHA wedge before installation. The iron tiles alternate with the scintillator tiles. The light guides to the photomultipliers are clearly visible.

CES

At the radial depth of $\sim 5.9 X_0$ in the CEM, i.e. close to the maximum in the longitudinal development of the electromagnetic showers, a proportional chamber named Shower Max Detector (CES, see figure 2.17), measures the local released ionisation projected in the two transverse directions.

The CES resolution is about 1 cm in z and about 1 mm in $r \cdot \phi$. During the Run 1 the CES information helped to reduce the fake electron trigger rate by a factor ~ 2 . A similar measurement is performed in the forward electromagnetic calorimeter by means of planes of crossed scintillator bars.

Plug Calorimeters

The plug calorimeters of Run II employ the same technology as the central ones and extend the coverage to $\eta \simeq 3.6$.

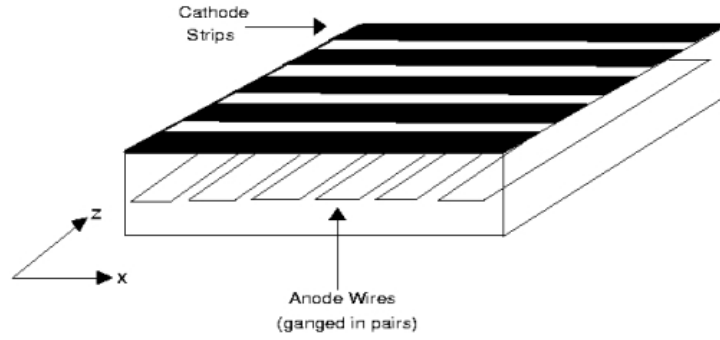


Figure 2.17: The CES detector in CEM. The cathode strips run in the x direction and the anode wires run in the z direction providing x and $(r \cdot \varphi)$ measurements

Each Plug Calorimeter is divided in 12 concentric η regions (figure 2.18), which are further segmented in azimuthal wedges (48 φ wedges at $|\eta| < 2.11$, 7.5° wide each, and 24 wedges at $|\eta| > 2.11$, 15° wide each) with transverse dimensions depending on depth in order to build an array of projective towers.

As in the central calorimeter, there is a front EM compartment and a rear hadronic compartment (PEM and PHA). The first has 4.5 mm thick lead tiles alternating with 4.0 mm thick scintillators for a total of 23 layers equivalent to $21 X_0$, the second has again 23 layers, but composed of 50.8 mm iron and 6 mm scintillator, for a total absorption length of $\sim 7.0 \Lambda_0$.

As in the central region, also PEM contains a shower maximum detector (PES) at $\sim 6 X_0$ depth. The PES consists of two layers of 200 scintillating bars each, oriented at crossed relative angles of 45° ($\pm 22.5^\circ$ with respect to the radial direction). The position of a shower on the transverse plane is measured with a resolution of $\sim 1 mm$.

In the Plug region the first scintillator layer is thicker and is read out separately to work as a pre-radiator (PPR). The Plug assembly is very compact, so that plug cracks are negligible.

The most important calorimeter parameters are given in table 2.2.

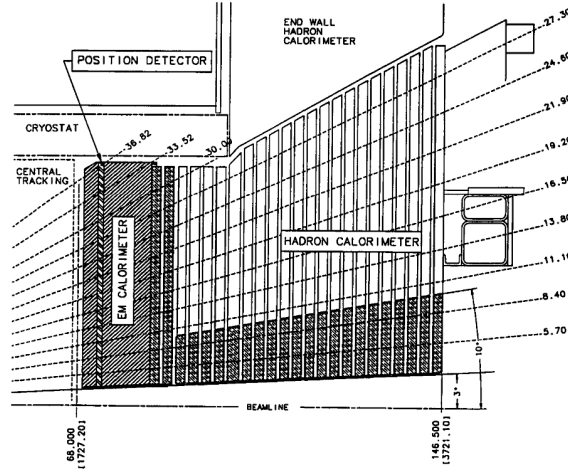


Figure 2.18: Plug Calorimeter (PEM and PHA) inserted in the Hadron End Wall calorimeter WHA and into the solenoid.

2.2.7 Muon Detectors

The muon detection system consists of several subdetectors, (CMU, CMP, CMX, IMU) as shown in figure 2.20, placed outside the hadron calorimeters. They are important as trigger elements as well as in the off-line analysis of muon events.

CMU

The Central Muon Chambers (CMU) [23] is a set of four layered drift chamber sandwiches housed on the back of wedges inside the central calorimeter shells covering the region $|\eta| < 0.6$. CMU is largely unchanged from Run 1, except for the fact that it operates now in proportional mode rather than in limited-streamer mode.

CMP

The Central Muon uPgrade (CMP) consists of a 4-layer sandwich of wire chamber operated in proportional mode covering most of the $|\eta| < 0.6$ region where it overlaps with CMU (see figure 2.20). Unlike mostly of the CDF

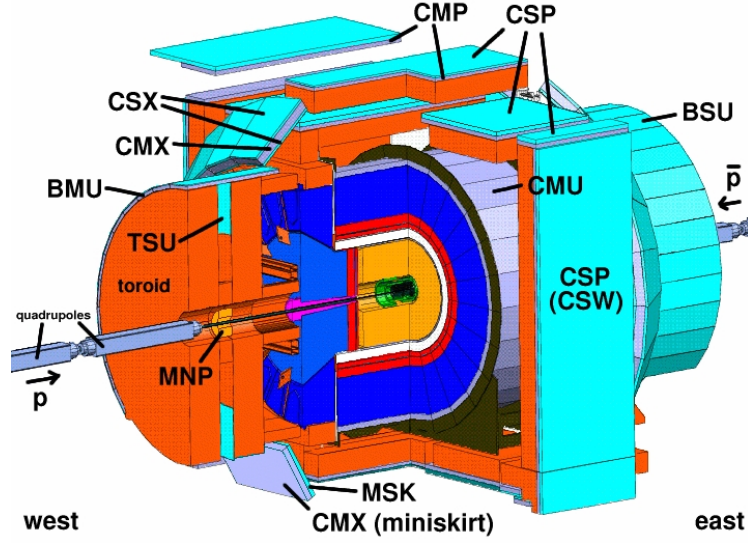


Figure 2.19: The CDF muon system.

components, this subdetector is not cylindrically-shaped but box-like, because CMP uses the magnet return yoke steel as an absorber, along with some additional pieces of steel to fill gaps in the yoke. On the outer surface of CMP a scintillator layer, the Central Scintillator Upgrade (CSP), measures the muon traversal time.

The system CMU/CMP, which is called CMUP, detects muons having a minimum energy of $\sim 1.4 \text{ GeV}$.

CMX

The muon extension CMX is a large system of drift chambers-scintillator sandwiches arranged in two truncated conical arches detached from the main CDF detector to cover the region $0.6 < |\eta| < 1.0$. Due to main detector frame structure, some regions of this subdetector are characterized by their peculiar geometry, as shown in figure 2.21.

IMU

Muons in a more forward region at $1.0 < |\eta| < 1.5$ are detected by the Intermediate Muon Extensions (IMU) on the back of the Plug Calorimeters.

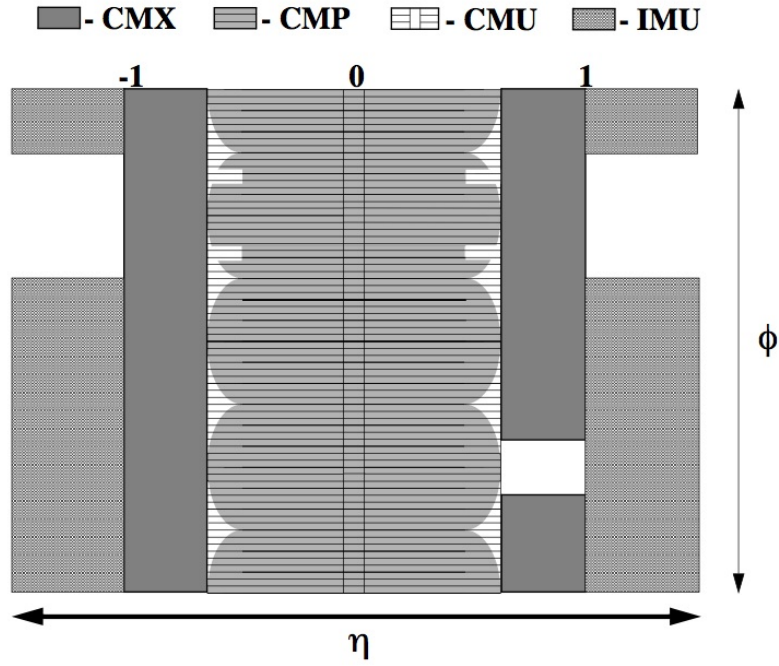


Figure 2.20: The η/ϕ coverage of the muon system. The shape is irregular because of the obstruction by systems such as cryo pipes or structural elements.

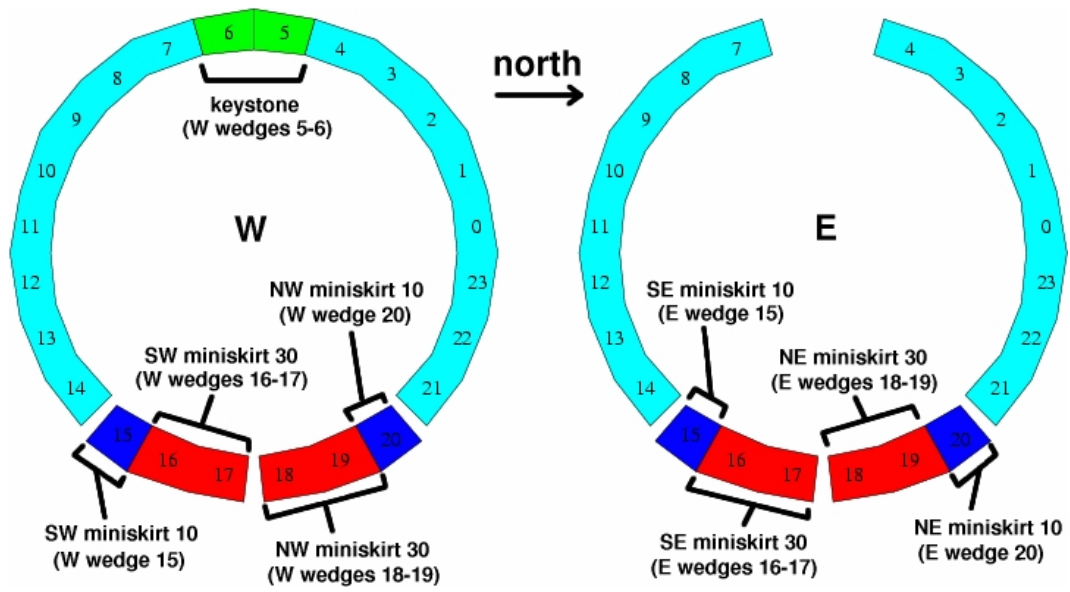


Figure 2.21: CMX subdetector scheme. The *arches*, *miniskirt* and *keystone* zones are shown.

A telescope of two toroidal iron shields IMU from the hot hadron flux in this angular region.

2.2.8 Cherenkov Luminosity Counter: CLC

In Run II CDF measures the collider luminosity with a coincidence between two arrays of Cherenkov counters, the CLC, placed around the beam pipes on the two detector sides. The counters measure the average number μ of interactions per bunch crossing. The luminosity is derived from the known average number of secondaries and inelastic cross section over the CLC angular coverage:

$$\mathcal{L}_{inst} = \frac{\mu f_{bunch}}{\sigma_{p\bar{p}}} \quad (2.5)$$

In this expression, $\sigma_{p\bar{p}}$ is the value of the inelastic $p\bar{p}$ cross section at 1.96 TeV and f_{bunch} is the rate of bunch crossing. This method measures the luminosity with about the 6% systematic uncertainty.

Each CLC module contains 48 gas Cherenkov counters of conical shape projecting to the nominal interaction point, organized in concentric layers. It works on the principle that light produced by any particle originated at the collision point is collected with a much higher efficiency than for background stray particles. The CLC signal is thus approximately proportional to the number of traversing particles produced in the collision.

2.2.9 Forward Detectors

The CDF Forward Detectors (whose scheme is shown in figure 2.22) include the Roman Pots detectors (RPS), beam shower counters (BSC) and two forward Mini Plug Calorimeters (MP). These detectors enhance the CDF sensitivity to production processes where the primary beam particles scatter inelastically in large impact parameter interactions (*pomeron exchange interactions*).

The Tevatron complex allowed to arrange a proper spectrometer making

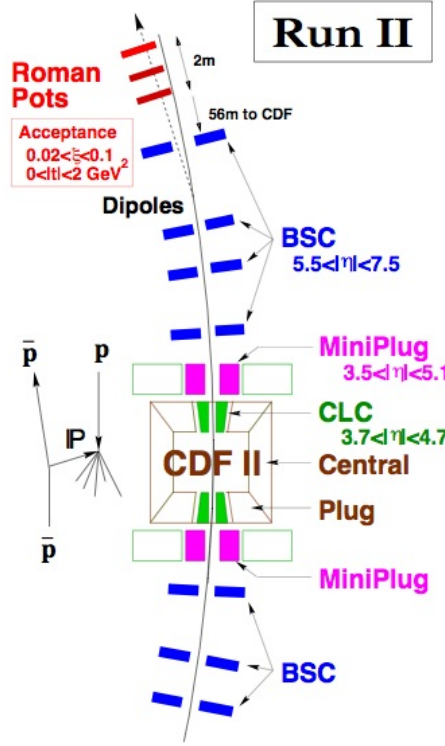


Figure 2.22: The forward detectors system in CDF, as arranged for Run II.

use of the Tevatron bending magnets only on the antiproton side. On this side, at appropriate locations, scintillating fiber hodoscopes inside three RPS measure the momentum of the inelastically scattered antiproton. Only the direction of the scattered proton is measured on the opposite side. The BSC counters at $5.5 < |\eta| < 7.5$ measure the rate of charged particles around the scattered primaries.

The MiniPlugs calorimeters at $3.5 < |\eta| < 5.1$ measure the very forward energy flow. MiniPlugs are a single compartment integrating calorimeter, consisting of alternate layers of lead and liquid scintillator read by longitudinal wavelength shifting fibers (WLS) pointing to the interaction vertex. Although the miniplug is not physically split into projective towers, its response can be split into solid angle bins in the off-line analysis. The MiniPlug energy resolution is about $\frac{\sigma}{E} = \frac{18\%}{\sqrt{E}}$ for single electrons.

2.2.10 Trigger System

The trigger system selects and stores interesting events for physics. Since the total inelastic cross section for $p\bar{p}$ is about $\sigma_{inel} \simeq 70 \text{ mb}$, and the collider luminosity can reach the $L = 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$, the event rate can reach several 10^6 events per second, while storage can only register events at 120 Hz maximum. If an appropriate selection is made at trigger level, a major loss of physics information can be avoided because the rate of really interesting events is very low.

As an example, one may note that the top quark production cross section is about 6 pb , which is $\sim 10^{-10}$ of the total inelastic cross section. A top event is then expected every 10 billion interactions. The top event rate would be one event in 10000 seconds at a luminosity of $10^{32} \text{ cm}^{-2} \text{ s}^{-1}$. If non-top events were efficiently rejected by the trigger the data logging power of CDF would be much more than adequate to collect all Top quark production events.

Level 1 Trigger

The CDF Trigger system is organized in three levels. The front-end electronics of all detectors is interfaced to a synchronous pipeline where up to 42 subsequent events can be stored for 5544 ns while the hardware is making a decisions. If by this time no decision is made, the event is lost. By doing so, since all the Level 1 decisions are made within about 4 μs , no dead time is caused by the trigger at Level 1.

The Level 1 rejection factor is about 150 and the maximum output rate is about 30 kHz .

The Level 1 accepts are generated by:

- XFT (eXtremely fast Tracker), which reconstructs tracks on the transverse plane of the COT. These tracks can be propagated to the calorimeter and to the muon chambers to contribute to electron or muon triggers at higher level;

- the calorimeter trigger, signaling large energy releases in the electromagnetic or hadronic cells as seeds of possible large energy electrons or hadron jets to be selected at higher level, and the total missing transverse energy;
- the Muon trigger, matching XFT tracks to stubs in the muon chambers.

Level 2 Trigger

Events accepted at Level 1 output are sent to 4 asynchronous buffers and further analyzed by a second set of hardware processors at Level 2. In Level 2 events are stored in the buffers until an individual decision is made. Because of the limited number of buffers some deadtime can be generated. In normal running conditions the Level 2 deadtime can be limited to less than 5%.

The main Level 2 trigger operations are:

- add the energy deposited in adjacent towers to the Level 1 seeds, as an approximate measure of an electron or jet energy;
- combine calorimeter and shower max detector (CES) information to improve the electron signature;
- signal tracks with large impact parameter by means of the Silicon Vertex Tracker (SVT). This important function permits to trigger on secondary vertexes from decay of long-lived beauty hadrons.

The Level 2 accept rate is up to 1 kHz and the rejection factor is about 150.

Level 3 Trigger

The Level 3 trigger is a software trigger, run on a Linux PC farm where all events are fully reconstructed using C++ codes and object-oriented techniques.

Events coming from Level 2 are addressed to the Event Builder (EVB), which associates information on the same event from different detector parts. The events can thus be fully reconstructed in the Level 3 processors. The final decision to accept an event is made on the basis of its features of interest (large

E_t leptons, large missing E_t , large energy jets and a combination of such) for a physics process under study, as defined by trigger tables containing up to about 150 entries. Events exit Level 3 at a rate up to about 100 Hz and are permanently stored on tapes for further off-line analysis. Each stored event is about 250 kB large on tape.

2.2.11 Online Monitoring

During data taking the quality of collected events is continuously monitored. A fraction of about $\sim 1\%$ of the on-line reconstructed events are copied into a computer center adjacent to the Control Room of the experiment where a number of "consumer programs" generate significant plots (individual trigger rates, subdetector occupancy, readout errors, hot and dead detector channels, etc.) for the CDF shift crew to evaluate and intervene in case of problems.

2.2.12 Data Processing

Raw data stored on tapes are split into streams according to trigger sets tuned to a specific physics process and are stored on fast-access disks. Data are fully reconstructed again by the CDF off-line code (*production*) using the best detector calibrations and reconstruction codes available at the time. Occasionally, if more detailed calibrations or significantly improved codes become available, data are re-processed. Re-processing is an heavy computer time-consuming operation which is performed only when significant gain in reconstructed event quality is expected. For the analysis performed in the present work, the reconstruction code versions 6.1.4 is used.

Chapter 3

Neural Network based algorithm for Heavy Flavour jets identification

A b quark produced in a $p\bar{p}$ collision emerges from the interaction region, after the hadronization process, as a B hadron. The B hadrons are bound states of a b quark with one or two more quarks and are characterized by high mass, long life-time (typically 1.5 ps) and high branching ratio for semileptonic decays.

A b -tagger is an algorithm that exploits these features to identify $high - P_T$ jets originated by b quarks. In this chapter, after a brief review of the b -tagging techniques used by the CDF collaboration, the novel Neural Network based algorithm for Heavy Flavour jets identification is described in detail.

3.1 b -tagging at CDF II

The CDF collaboration developed three main b -tagging techniques [24] for $high - P_T$ jets: SecVtx, JetProbability and Soft Lepton Tagger.

- *SecVtx*: this tagger, using the tracks inside the jet cone, tries to reconstruct a secondary vertex, i.e. a vertex displaced from the primary interaction one. Since also C hadrons can provide displaced vertexes

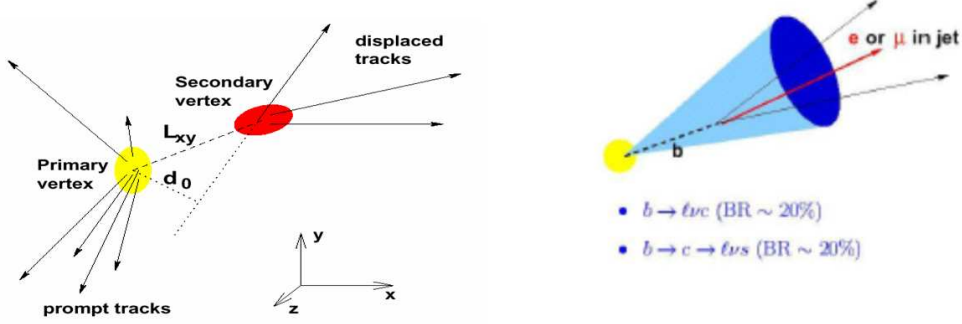


Figure 3.1: Left panel: scheme of a secondary vertex reconstructed by displaced tracks. Right panel: semileptonic decay of a b quark inside the jet cone.

inside the jet cone this tagger evaluate also a “vertex mass” to increase the capability to distinguish b -like from c -like events. The tagger’s performances are also limited by the resolution for primary and secondary vertex reconstruction.

- *JetProbability* : this tagger is based on the possibility of assigning to each track a “probability” of coming from the primary vertex based on its impact parameter signed with respect to the jet axis. Combining the probability for the well-identified tracks in a jet it is possible to evaluate a probability for the jet itself (“JetProbability”) to be composed by particles consistent with coming from the primary vertex. The distribution of this probability is, by construction, flat for jets originated by *light* quarks, and is peaked at small values for b and c jets. The jet is “tagged” if the value of the JetProbability output is less than a given threshold (values typically used are 0.01 and 0.05).
- *Soft Lepton Tagger* : this tagger search inside the jet cone for *low* – P_T leptons coming from a semileptonic decay of the B hadron. The efficiency of this technique is limited by the branching ratio ($\sim 20\%$) of the semileptonic decay and by the lepton identification efficiency. The

electron version (SLT_e) is under development, while the muon version (SLT_μ) is currently used into the CDF II analysis.

The taggers are characterized by their efficiency in b quark identification and by their “mistag rate”, i.e. the efficiency in the mis-identification of a *light* quark as a b quark. These efficiencies and mistag rates are evaluated in a combination of Monte Carlo and in suitable data control samples. Differences between data and MC are compensated by suitable “scale factors”.

3.2 Neural Network tagger

All the techniques reported in the previous paragraph use only a limited part of the information collected by the detector for a given event to identify *high* – P_T jets originated by a b quark.

A new tagger has been developed in order to exploit as much information as possible to identify the flavour of the quarks originating a jet.

This tool presents two main characteristics:

- a new vertexing algorithm able to reconstruct not only a single displaced secondary vertex, but several vertexes inside the jet cone, to take advantage of the frequent decay patterns with secondary and tertiary vertexes (e.g. $b \rightarrow c \rightarrow \text{light}$);
- a chain of Neural Networks, suitable to combine all the available information in a single discriminant.

The scheme of the tagger architecture is reported in figure 3.2. The new vertexing algorithm, starting from primary vertex and tracks information, tries to reconstruct all the vertexes inside the jet cone. The list of vertexes produced, together with tracks information is provided to a first Neural Network trained to discriminate between heavy-flavour and light-flavour vertexes.

A second Neural Network uses the output of the vertexes NN and tracks information to recover the tracks not associated to a vertex. General jet information, the response of the other taggers (SecVtx, JetProbability and SLT_μ), the

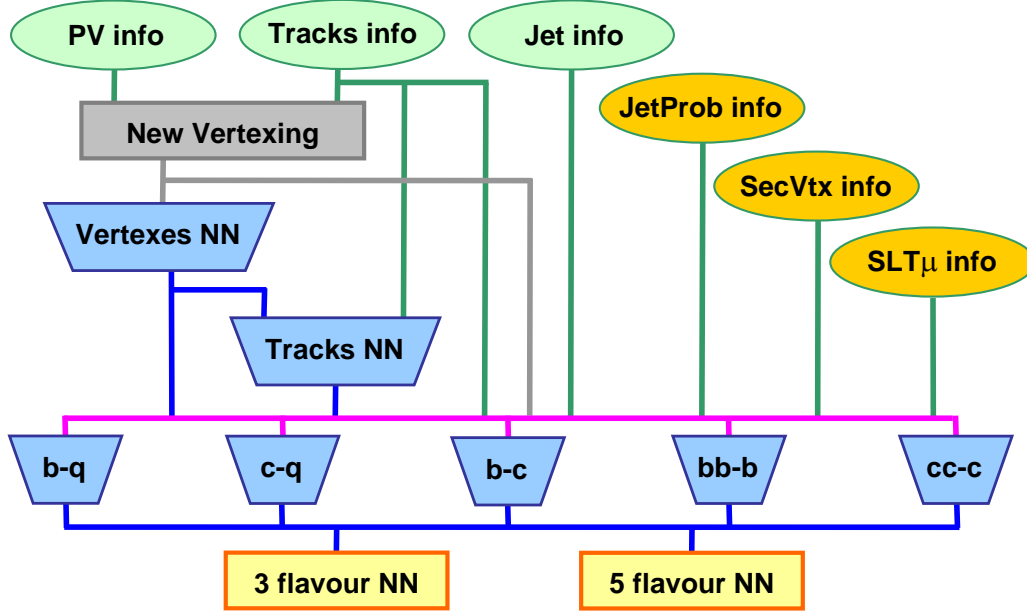


Figure 3.2: Scheme of the Neural Network based tagger.

tracks information, the list of the vertexes found and the output of the vertexes NN and of the tracks NN are the inputs to a set of Neural Networks trained to separate quark flavours. Five main jet categories has been identified: jets originated by b quark, c quark or a *light* quark (u, d, s), and jets containing two heavy quarks ($b\bar{b}$ or $c\bar{c}$) inside the cone. The implemented jet Neural Networks are configured to discriminate between 2-flavour hypothesis: $b - q$, $c - q$, $b - c$, $b\bar{b} - b$, $c\bar{c} - c$. Two final Neural Network process the outputs of these flavour-discriminating Neural Networks and provide a 3-flavour discriminant ($b - c - light$) and a 5-flavour discriminant ($b - b\bar{b} - c - c\bar{c} - light$).

For the implementation of all the Neural Networks, except the final discriminants, the NeuroBayes [25] package was used. This provide an efficient tool for the ordering of the input variables according to their significance and a Neural Network that can be trained to discriminate between two categories of events. The final output of the chain is instead provided by TMultiLayer-Perceptron Neural Networks, available in the ROOT package [26], which can classify more than two categories of events with a single output.

	Good $r - \phi$	Good $r - z$
track P_T	$> 0.5 \text{ GeV}/c$	
corrected $r - \phi$ impact parameter $ D0 $	$< 0.15 \text{ cm}$	-
corrected $r - z$ impact parameter $ D0 $	-	$< 1.0 \text{ cm}$
number of $r - \phi$ hit layer (SVX-II)	$\geq 3 \text{ \& } \leq 5$	-
total number of COT axial hits	≥ 20	
total number of COT stereo hits	≥ 17	
$ Z_{trk} - Z_{pv} $	$< 5 \text{ cm}$	-

Table 3.1: Selection criteria for tracks used in the evaluation of the JetProbability tagger.

A jet is defined *identifiable* – i.e. the tagger can be evaluated – if it has at least either one vertex reconstructed by the new vertexing algorithm (this vertex does not need to be HF-like) or 2 tracks marked good by the JetProbability tagger with criteria reported in table 3.1.

3.2.1 Training samples

The Monte Carlo training samples are composed by 50% jets from $Z(\rightarrow e^+e^-)$ + parton events and 50% jets from $t\bar{t}$ events, and are simulated using Pythia.

The jets are reconstructed using the JETCLU algorithm, with a $R = \sqrt{|\Delta\eta|^2 + |\Delta\phi|^2} = 0.4$ cone radius. The jet flavour is determined considering the flavour of any quark lying inside the jet cone. The selected jets are required to be separated from the electrons of the Z decay by at least $\delta R = 0.4$. Furthermore the training samples are constructed to have a uniform distribution in jet transverse energy (E_T) between 15 GeV and 200 GeV. In this way the jet E_T can be included in the Neural Network input variables (several discriminating variables, e.g. L_{xy} , have a dependence on the B hadron boost and therefore on the jet E_T) without biasing the training to a particular jet E_T distribution.

3.2.2 Vertexing algorithm

The new vertexing algorithm provides for each jet a list of identified vertexes, using as input the tracks lying inside the jet cone and the primary vertex of the event reconstructed with the standard CDF II algorithm [27]. The selected tracks must satisfy $p_T \geq 1 \text{ GeV}/c$, $z_0 < 1 \text{ cm}$ and $d_0 < 0.15 \text{ cm}$, where z_0 and d_0 are the projections, respectively along the beam line and on the transverse plane, of the minimum distance of the track from the primary vertex. Tracks reconstructed using exclusively COT or silicon hits, or identified as originating from photon conversion are rejected, while tracks from K_S or Λ decay are kept, since b and c hadrons have significant decay ratios into these states. All vertex fits are performed by the standard CDF II algorithm CTVMFT [27].

The following procedure is applied to each pair of tracks:

- fit a vertex (seed vertex);
- if $\chi^2 > \chi_{seed}^2$ reject the seed vertex and go to the next pair of tracks;
- find the track (not already in the vertex) which gives smallest $\delta\chi^2$ when added to vertex; if $\delta\chi^2 < \chi_{add}^2$ add to the vertex and refit vertex; repeat for all the tracks;
- find the track which gives the largest $\delta\chi^2$ to the vertex; if $\delta\chi^2 > \chi_{add}^2$ remove from the vertex and refit vertex; repeat for all the tracks assigned to the vertex;
- add vertex to the list of vertexes.

After an optimization procedure on a Pythia Monte Carlo sample, the values for the thresholds χ_{seed}^2 and χ_{add}^2 were chosen to be respectively 6 and 24.

At this level it is possible that a track may be used in the reconstruction of more than one vertex. The following procedure is then applied to ensure unique association between tracks and vertexes:

- first considering only vertexes with 3 or more tracks

- calculate the χ^2 contribution of each track to each vertex in which it is included;
- if the largest χ^2 contribution is a shared track remove from vertex and refit vertex;
- repeat the procedure while there are still shared tracks;
- then considering only vertexes reconstructed with 2 tracks (even if produced from the previous step)
 - find the vertex with lowest χ^2 and shared tracks;
 - remove 2-track vertexes which have tracks in common with this vertex;
 - repeat the procedure while there are still shared tracks;

3.2.3 Vertexes classification

Since there is no selection on impact parameter of seed tracks, the vertexes found by the algorithm described in the previous paragraph can be either the primary vertex reconstructed by prompt tracks in the jet, a true secondary vertex reconstructed using the decay tracks of a long-lived particle, a fake displaced vertex (including mismeasured tracks, for example), or may be a “mixed” vertex containing both prompt and secondary tracks.

Several vertex observables were considered:

- invariant mass of tracks in vertex;
- vertex pseudo $c\tau = L_{xy} \cdot \frac{M^{vtx}}{P_T^{vtx}}$;
- separation between the vertex position and the primary vertex (PV), projected along the jet axis, in both 2 and 3 dimensions;
- significance of the vertex–PV separation;
- $\Delta\phi$ and 3d-angle between vertex momentum and the vector between vertex and the primary vertex;
- vertex p_T with respect to the jet direction;
- number of tracks in the vertex;
- charge of the vertex.

rank	variable	significance
1	vertex pseudo $c - \tau$	118.0
2	2 nd track d_0 significance	41.2
3	angle between vertex mom. and displ.	25.3
4	L_{xyz} significance	12.2
5	invariant mass	6.5
6	L_{xy}	4.3
7	1 st track d_0 significance	4.2
8	2 nd track d_0	3.4
9	$\delta\phi$ between vertex mom. and displ.	3.0
10	χ^2 of vertex fit	3.0
11	L_{xyz}	3.0
12	L_{xy} significance	1.4
13	charge of vertex	1.1
14	# tracks in vertex	0.7
15	1 st track d_0	0.2

Table 3.2: Rank of input variables to the vertexes Neural Network, from NeuroBayes. The first 5 variables are used in the final network.

Many of these observables provide some separation between the different vertex samples, and several sets of observables are correlated. The list of the observables considered is reported in table 3.2 according to their separation power ranking evaluated by NeuroBayes package. The observables with the highest significance are used as input variables to the “vertexes Neural Network” trained to distinguish primary, fake and mixed vertexes from those whose tracks all come from b and c hadron decays. In particular the first five observables are used, because training the network with a longer list did not give any appreciable improvement in performance.

The jets are classified as b , c , $b\bar{b}$, $c\bar{c}$ or *light* according to the number of quarks of the relevant flavour found inside the jet cone. The vertexes identified inside each jet are classified as a true $b(c)$ vertexes if all the associated tracks

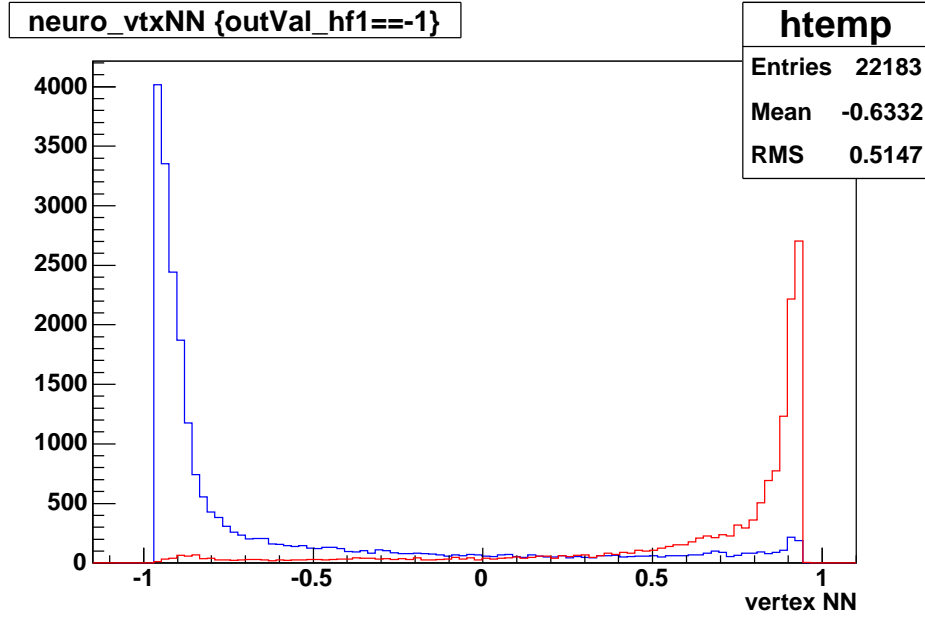


Figure 3.3: Distribution of the vertexes Neural Network output for a test sample with the same statistics and composition used in the training; the HF vertexes are in red and the non-HF vertexes are in blue.

are matched to a $b(c)$ daughter particle, while if a vertex contains tracks not from $b(c)$ decay, it is labeled as a non-HF vertex. For the training procedure several different samples are used:

1. true HF vertexes in matched b jets;
2. true HF vertexes in matched c jets;
3. all vertexes from *light* quark jets;
4. non-HF vertexes in matched b jets;
5. non-HF vertexes in matched c jets.

Each sample contains around 3700 vertexes each from $Z + p$ and $t\bar{t}$ events weighted according to their classification. To the vertexes in classes 1, 2 and 3 is assigned a weight of 1 in the training; classes 4 and 5 may contain mixed vertexes, where some tracks are from HF, others not, so these enter the training with a lower weight (0.5).

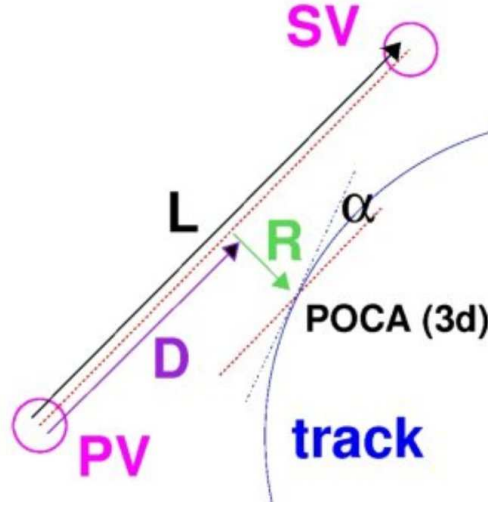


Figure 3.4: Definitions of the observables used in track selection. PV is the primary vertex; SV is a secondary vertex; POCA is the point of closest approach of the track to the segment connecting PV to SV.

The network is implemented with 10 hidden nodes, and is trained to distinguish between pure HF vertexes (output = +1) and the others (output = -1). The distribution of the Neural Network output is shown in figure 3.3.

3.2.4 Unvertexed tracks

In the cascade decay of b hadrons some tracks can be not associated to any identified vertex. This stage of the algorithm is devoted to recover the information about the jet flavour contained in this kind of tracks. The various discrimination observables useful to classify the unvertexed tracks are defined in figure 3.4.

Each unvertexed track is compared to each identified vertex according to the following preselection intended to clean up tracks far from the secondary vertex, and to consider only comparisons to HF-like vertexes:

- vertexes $NN > 0.0$
- $R < 1 \text{ cm}$
- $D < 5 \text{ cm}$

rank	variable	significance
1	track d_0 significance	78.6
2	D/L	35.0
3	vertexes NN	31.7
4	α	18.1
5	D	6.9
6	track p_T	6.3
7	R significance	4.4
8	track d_0	2.2
9	R	0.3

Table 3.3: Rank of input variables to the track Neural Network, from NeuroBayes. The first 5 variables were used in the final network.

- $|D/L| < 5$

According to this criteria more than one track-vertex combination can be considered for each unvertexed track, both in the training and in tagger evaluation. The following variables were considered as possible discriminating inputs to the tracks Neural Networks:

- the angle α
- track p_T
- track d_0 and significance (signed with respect to the jet axis)
- R and its significance
- D
- D/L
- NN_{vtx} of vertex to which track is compared

The ranking of the observables produced by the NeuroBayes package is shown in table 3.3. The first 5 variables in the list are used as input to the tracks Neural Network; using more inputs did not appreciably improve the network performance.

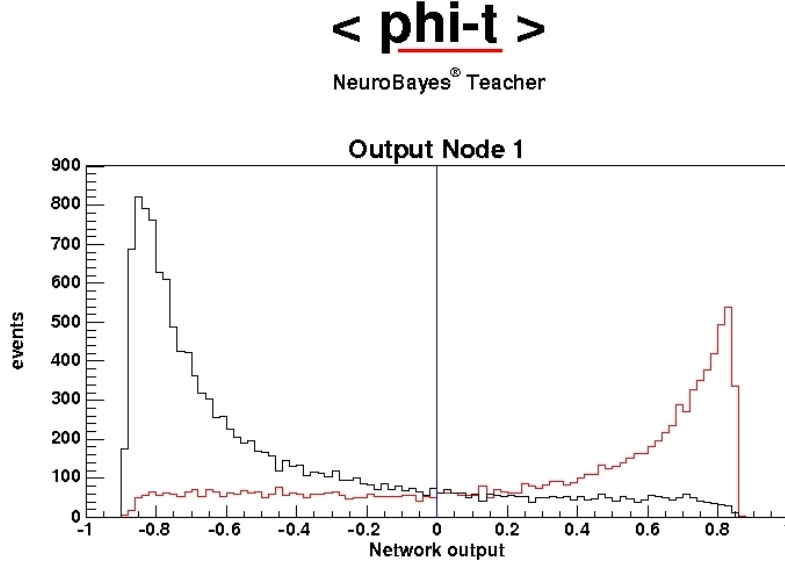


Figure 3.5: Output of the track Neural Network for a test sample with the same statistics and composition used in the training. The distribution of the HF unvertexed tracks is in red, while the distribution for the non-HF unvertexed tracks is in black.

The training samples have around 3700 track–vertex combinations in each of $Z+p$ and $t\bar{t}$ events in the following categories:

1. true HF track compared to a true HF vertex in matched b or c jets (output=+1);
2. all track–vertex combinations in *light* quark jets (-1);
3. non HF track to non HF vertex in matched b or c jets (-1);
4. non HF track to true HF vertex in matched b or c jets (-1);
5. true HF track to non HF vertex in matched b or c jets (-1).

The network is implemented with 10 hidden nodes, and is trained to distinguish between HF unvertexed tracks (output = +1) and the others (output = -1). The distribution of the Neural Network output is shown in figure 3.5.

3.2.5 Jets discrimination

Information from vertexes and tracks with a Neural Network score greater than 0 are combined with jet observables to separate various types of jets: *light* jets, jets containing a *b* or a *c* quark and jets containing two *b* or *c* quarks. Several kinds of observables are potentially useful for this task:

- number of selected vertexes (i.e. vertexes with $NN > 0.0$);
- number of selected unvertexed tracks (i.e. tracks with $NN > 0.0$);
- invariant mass of all selected vertexes & tracks;
- characteristics of the vertexes with the first- and second-highest NN output;
- $\sum_{selected\ tracks} p_T / \sum_{all\ jet\ tracks} p_T$;
- largest $\Delta R = \sqrt{|\Delta\eta|^2 + |\Delta\phi|^2}$ between any 2 selected vertexes;
- p_T with respect to the jet axis of the muons identified by the *Soft Lepton Tagger*;
- information from other taggers (*SecVtx*, *JetProbability*).

The jet E_T does not provide any discrimination when used alone, since all training samples have an identical, flat E_T distribution between 15 and 200 GeV. However, in conjunction with other variables, it may provide useful discriminating information in cases where these variables are correlated with the jet E_T , so we consider it as a possible input variable.

The various NeuroBayes NNs are trained to separate between the following jet flavours:

- *b jets* vs. *light jets*
- *c jets* vs. *light jets*
- *b jets* vs. *c jets*
- $b\bar{b}$ jets vs. *b jets*
- $c\bar{c}$ jets vs. *c jets*

All the variable which have a separation significance (as given by NeuroBayes) of at least 5 in one or more of the above networks are used as input to all the 5 networks. The input list is composed by the following 16 variables:

- JetProbability (when calculated with at least 2 tracks);
- pseudo $c - \tau$ of vertex with highest NN;
- loose SecVtx tag;
- mass of SecVtx vertex;
- number of muons identified by SLT $_{\mu}$;
- p_T of (highest p_T) muon w.r.t. jet axis;
- $\sum_{selected\ tracks} p_T / \sum_{all\ jet\ tracks} p_T$;
- L_{xyz} significance of vertex with highest NN;
- invariant mass of vertex with highest NN;
- invariant mass of selected tracks (unvertexed tracks with NN > 0.0 plus tracks belonging to vertexes with NN > 0.0);
- number of selected tracks;
- L_{xyz} of vertex with highest NN;
- number of good tracks inside the jet cone ($p_T > 1.0\text{ GeV}/c$, $d_0 < 0.15\text{ cm}$, $z_0 < 1.0\text{ cm}$);
- $\sum_{all\ jet\ tracks} p_T$;
- L_{xyz} of vertex with second highest NN;
- jet E_T ;

The distributions of the networks outputs are shown in figures 3.6 and 3.7. There is some separation between all flavours: the best separation is between b and *light* jets (as expected), there is reasonable separation between b - c and c - q jets, while the separation between $b\bar{b} - b$ and $c\bar{c} - c$ is more limited.

3.2.6 Combined Networks

The last level of networks used by the new tagger is implemented by ROOT TMultiLayerPerceptron. This package, contrary to NeuroBayes, can be trained

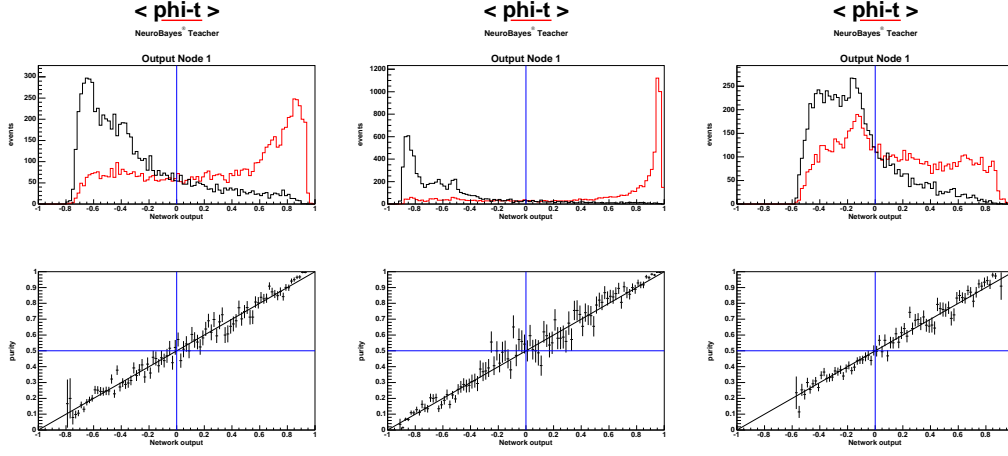


Figure 3.6: Output distributions for the 2-flavour Neural Networks are shown in the upper row. The corresponding purity distributions are reported in the lower row. Left panel: b (red) - c (black). Central panel: b (red) - q (black). Right panel: c (red) - q (black).

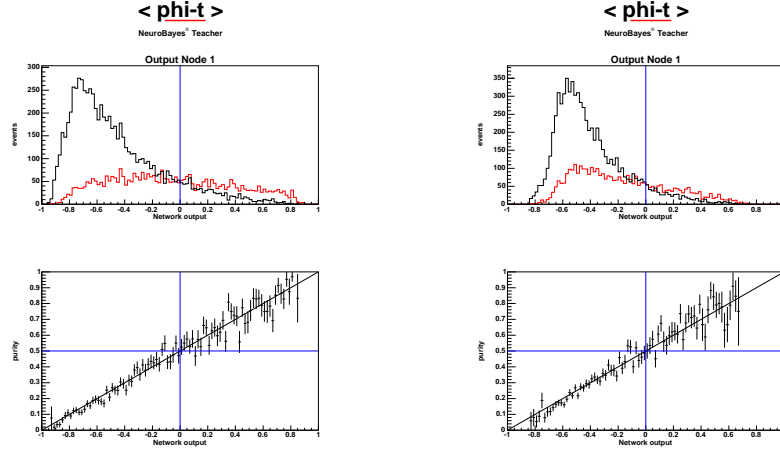


Figure 3.7: Output distributions for the 2-flavour Neural Networks are shown in the upper row. The corresponding purity distributions are reported in the lower row. Left panel: $b\bar{b}$ (red) - b (black). Right panel: $c\bar{c}$ (red) - c (black).

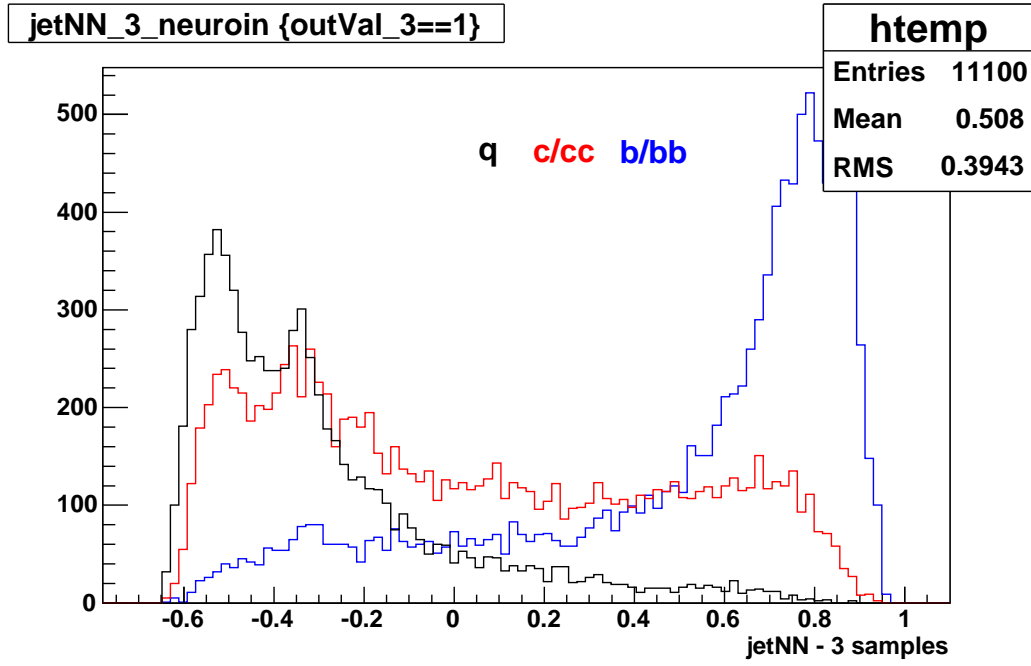


Figure 3.8: Output distribution for the 3-flavour jet Neural Network.

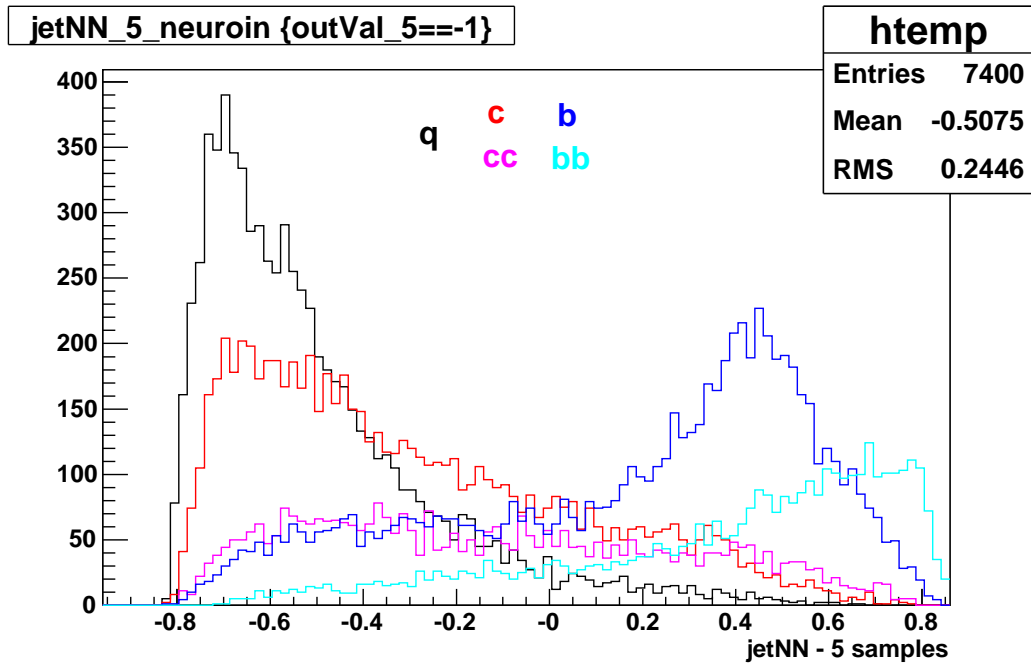


Figure 3.9: Output distribution for the 5-flavour jet Neural Network.

to separate more than two samples in a single output node. These final networks were trained to separate three and five jet categories respectively:

$$\textbf{3-flavour} \quad : \quad b/b\bar{b} \ (1) - c/c\bar{c} \ (0) - \textit{light} \ (-1)$$

$$\textbf{5-flavour} \quad : \quad b\bar{b} \ (1) - b \ (0.5) - c\bar{c} \ (0) - c \ (-0.5) - \textit{light} \ (-1)$$

The output value used for the training procedure is reported in parenthesis.

The outputs of the five NeuroBayes 2-sample networks are provided as inputs to these two multi-sample networks. This system of networks can be thought of as a network with 2 hidden layers: the second layer is not hidden, however, and its nodes have physical meaning since they separate pairs of flavours in an optimal way. The same sample of jets from $Z + p$ and $t\bar{t}$ events, with flat distributions in jet E_T , was used to train these networks. The distributions of the combined network outputs are shown in figures 3.8 and 3.9.

Given the reduced separation between single and double heavy flavour obtained by the $b - b\bar{b}$ and $c - c\bar{c}$ networks, the 3-flavours NN provide a larger separation among the components, with respect to the 5-flavours one. For this work only the 3-flavours NN has been used.

The continuous distribution of the tagger output between -1 and 1 makes it suitable for two different kinds of application: *cut-based* or *shape-based*. In the first case all the jets with a NN output greater than a given threshold are labeled as b -jets (*tagged*), while all the other jets are identified as *non* - b -jets. In the shape-based approach, instead, all the NN spectra is used to assign to each jet a probability of belonging to a given category of events (e.g. b , c or *light*). This second approach allows for a more efficient utilization of the available statistics, and can provide the simultaneous measurement of different phenomena, but need also a better knowledge of the exact shape of the tagger output for every category of events.

In figure 3.10 is reported a comparison of the new 3-flavour tagger efficiency respect to the “SecVtx” b -tagger, for the cut-based application evaluated in a MC di-jet sample.

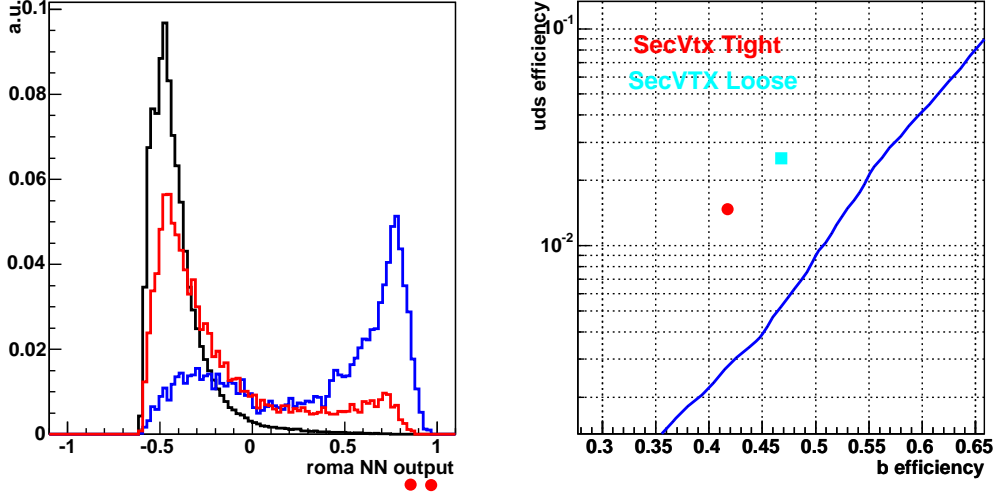


Figure 3.10: Left panel: distribution of the 3-flavour Neural Network for pure *b* (blue), *c* (red) and *light* (black) jets obtained in a di-jet MC sample. Right panel: comparison of the working points of the *Tight* and *Loose* versions of the SecVtx tagger with the corresponding curve obtained for the 3-flavour Neural Network tagger, varying the tagging cut.

3.3 Data - Monte Carlo comparison

The response of the new tagger on a generic jet data sample is compared to the corresponding Pythia Monte Carlo simulation to find out possible disagreements. The data sample is collected with the JET20 trigger which ask for at least one jet with $E_T > 20$ GeV. Di-jets events can be selected in both data and MC with the following requirements:

- *probe-jet* j_p : $E_T > 22$ GeV and $|\eta| < 1.5$;
- *away-jet* j_a : $E_T > 11$ GeV and $|\eta| < 1.5$;
- $\Delta\phi(j_p, j_a) > 2$;

The cut on the *probe-jet* E_T has been chosen higher than the trigger threshold to reduce possible trigger bias. The distribution of the 3-flavour network output for the leading jet is shown in figure 3.11, while in figures from 3.12 to 3.17 the corresponding distributions of the input variables to the five 2-flavour networks are reported.

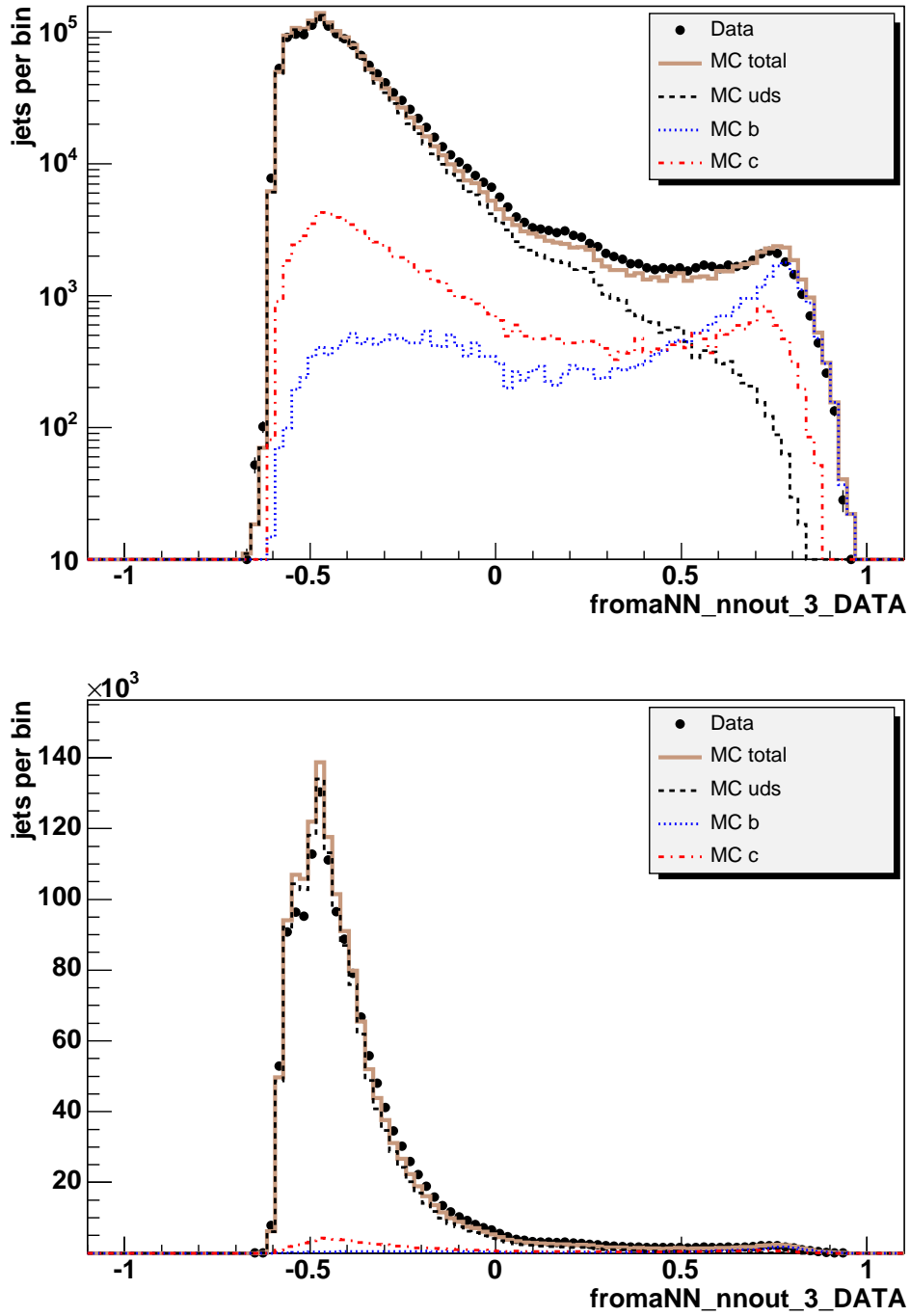


Figure 3.11: Output distribution of the 3-flavour Neural Network for JET20 data, compared with the corresponding Pythia Monte Carlo simulation in logarithmic (upper panel) and linear (lower panel) scale.

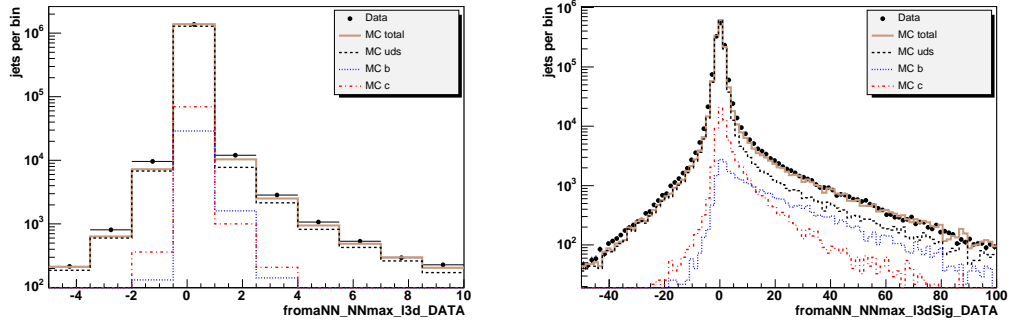


Figure 3.12: Left panel: L_{xyz} of the vertex with the highest Vertexes NN output. Right panel: L_{xyz} significance of the vertex with the highest Vertexes NN output.

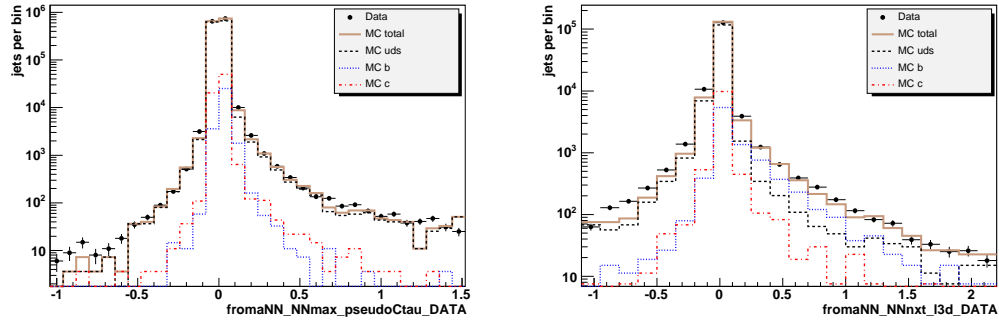


Figure 3.13: Left panel: $c\tau$ of the vertex with the highest Vertexes NN output. Right panel: L_{xyz} significance of the vertex with the second highest Vertexes NN output.

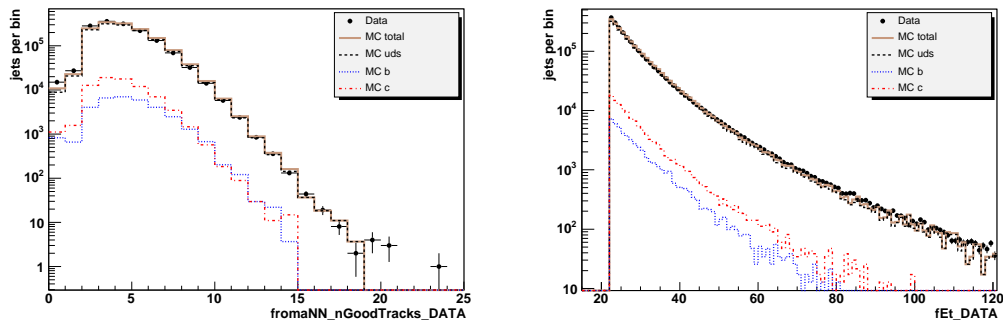


Figure 3.14: Left panel: number of tracks marked good for the JetProbability tagger. Right panel: jet transverse energy.

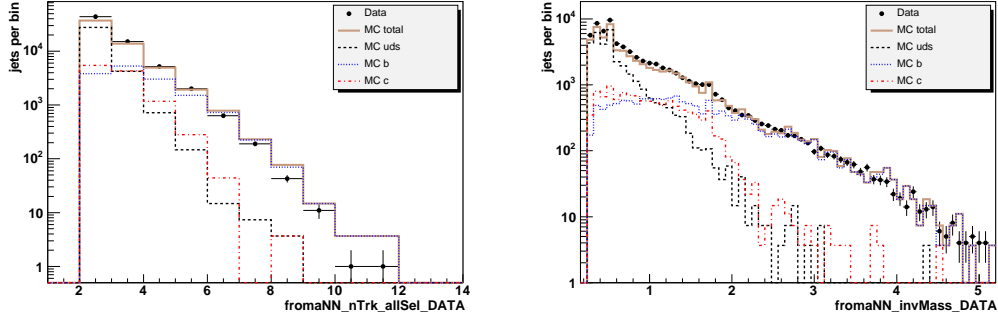


Figure 3.15: Left panel: number of tracks selected inside the jet cone. Right panel: invariant mass of selected tracks.

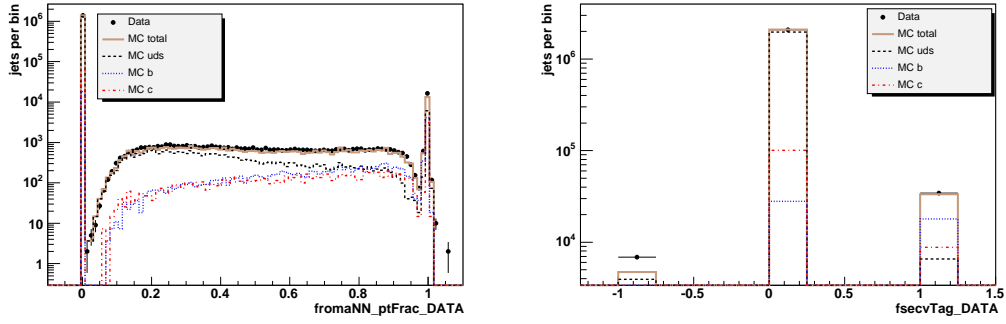


Figure 3.16: Left panel: $\sum p_T$ of the selected tracks divided by the $\sum p_T$ of all the tracks reconstructed inside the jet cone. Right panel: *Loose* SecVtx output.

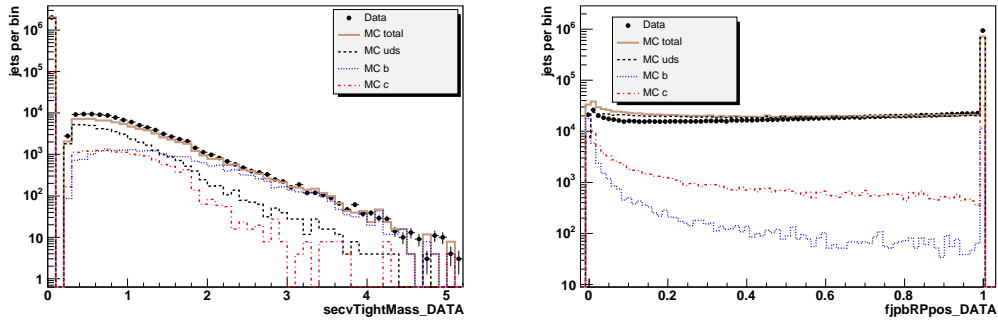


Figure 3.17: Left panel: mass of the secondary vertex reconstructed by the SecVtx tagger. Right panel: output of the JetProbability tagger.

Chapter 4

Calibration procedure

Every disagreement in the tagger response between data and Monte Carlo can affect any MC-based estimation. For the SecVtx tagger, which has a binary output, the standard procedure developed by CDF II collaboration to correct this kind of effect is based on the evaluation of *scale factors* for efficiency in suitable *b*-enriched control samples and calculation of a *mistag rate* matrix in *light* quarks dominated jet samples.

For the shape-based approach the whole shape of the normalized distribution of the tagger output is relevant, and so the correction itself has to be a function of the tagger output. Furthermore the correction has to be specific for each category of events among which the tagger is supposed to distinguish.

4.1 Calibration strategy

For the shape-based application to the fraction measurement, the correction of the Neural Network tagger for Monte Carlo events can be implemented as a weight function for each spectra of the tagger output for pure flavour MC samples (*b*-, *c*- and *light-template*).

Since this new tagger was developed with the aim of exploiting all the available information recorded by the detector, it is not possible to select a data sample highly enriched in heavy flavour content which is unbiased for the tagger. Therefore to extract the needed correction from a data sample it is nec-

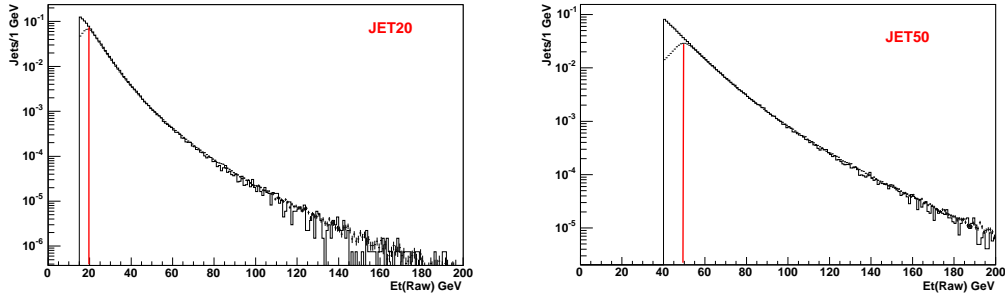


Figure 4.1: Transverse energy distribution for the *leading-jet* in JET20 (left panel) and JET50 (right panel) samples, compared with the corresponding Pythia Monte Carlo simulation. The red line indicates the trigger threshold.

essary to cope simultaneously with the determination of sample composition and with the extraction of the correction. A data-driven procedure has been developed to extract a *correction function* (cf) for each template, from a set of independent calibration samples - characterized by different heavy flavour content - selected from generic jet data. To reduce as much as possible any ambiguity in correction's assignment the number of the calibration samples has to be equal or greater than the number of the templates. Furthermore the calibration samples must have different flavour compositions in order to reduce the interplay between composition determination and correction extraction. The aim of the present analysis is to separate jets originated from three kinds of quarks (b , c and $light$), therefore the calibration procedure has been set to provide correction functions for the respective templates, even if this could, in principle, be applied to a wider list of jet categories, once a suitable set of calibration samples is provided.

4.2 Definition of calibration samples

All the calibration samples must be unbiased for the Neural Network tagger distribution, i.e. none of the variables used to evaluate the tagger itself can be exploited to enrich the sample in a particular flavour content. On the other hand, by construction all the variables not selected as input to the tagger, do

not provide a considerable separation among flavours. In order to build samples of jets with different Heavy Flavour content avoiding as much as possible any kind of bias, it is possible to exploit the flavour correlation of di-jet events, in which b or c quarks are pair produced at Tevatron.

In a di-jet event it is useful to define a *probe-jet* (j_p), i.e. the unbiased jet used to build the calibration samples, and an *away-jet* (j_a), i.e. the jet produced in association with the probe one and exploited to enrich the calibration sample in a particular flavour. Selecting different kind of away jets it is possible to change the b , c and *light* fractions in the probe jets samples:

- *b-enriched* (B) : j_a tagged by SecVtx and $LBLnn \geq 0.6$;
- *c-enriched* (C) : j_a tagged by SecVtx and $LBLnn < 0.6$;
- *light-enriched* (L) : j_a untagged by SecVtx.

The *light* flavour enriched sample is built requiring an away-jet with no secondary vertex reconstructed by SecVtx. On the contrary the b and c enriched samples ask for a secondary vertex inside the away-jet cone. Furthermore to distinguish between b and c jet, several conditions have been tested in the away-jet selection. The most effective one exploits a Neural Network based tagger developed by the Laurence Berkley Laboratory group: this network ($LBLnn$) works on the jet already tagged by SecVtx, producing a continuous output distributed between 0 and 1, and allows to increase the purity of the selected b sample. In figure 4.2 is reported the distribution of the $LBLnn$ for b , c and *light* away jets, and the predicted fractions in generic di-jet MC samples as a function of the cut on $LBLnn$ output. The threshold value for $b - c$ separation has been chosen in order to balance purity and number of selected events.

Due to the di-jet identification efficiency and to the away-jet tagging efficiency, the *light* fraction is dominant in all the calibration samples. However the three sets of fractions differ enough to assure a good statistical separation of the components with the available statistics. An example of the typical composition of the calibration samples, measured at the end of the calibration procedure, is reported in table 4.1 for the specific case of the $Z + jet$ analysis.

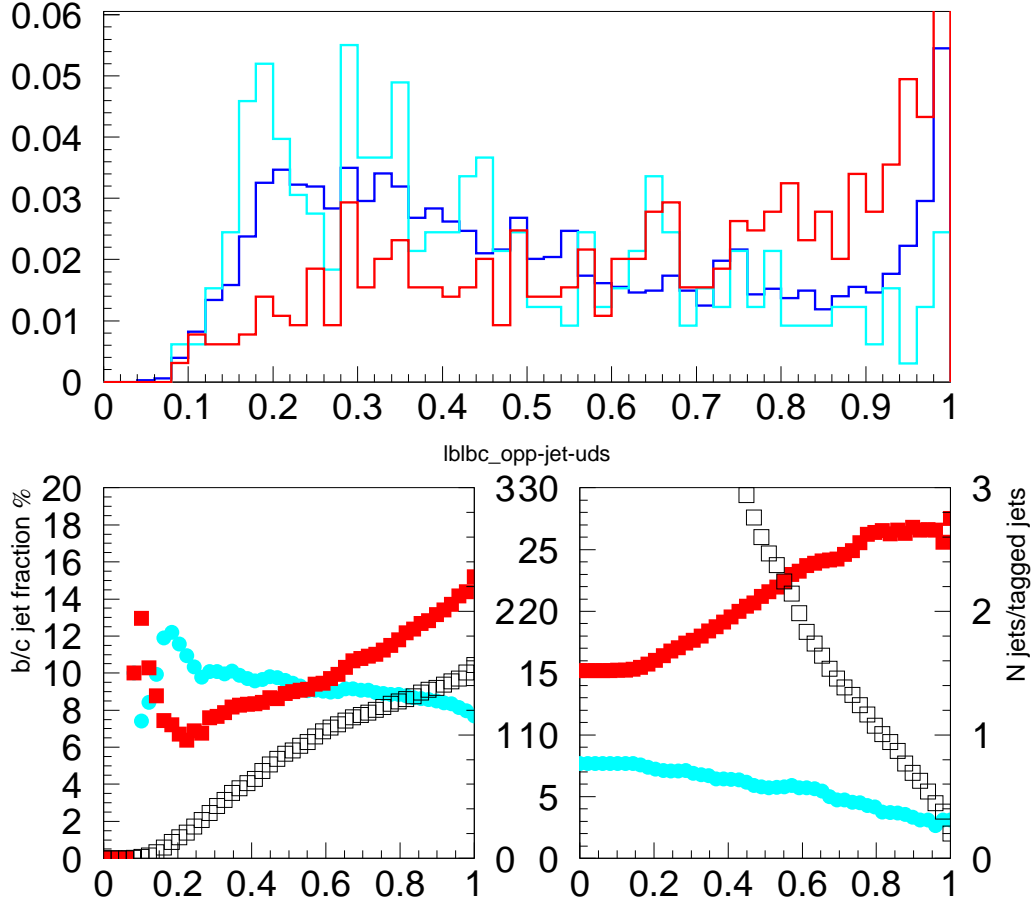


Figure 4.2: Upper Panel: Berkeley 2-flavor Neural Network b/c in SecVtx tight tagged *away-jet* for *light*, b and c -quark jets. Lower Right Panel: MC predicted *probe-jet* b (red) and c (cyan) content of *probe-jet* jets when selecting *away-jet* with Berkeley 2-flavor Neural Network b/c lower than cut value plotted on the x axis (C sample). Open black squares are number of data events relative to the nominal SecVtx selection (cut on the SecVtx mass $m_{SecVtx} > 1.5 \text{ GeV}/c^2$). Lower Left Panel: same for B sample reversing the value of the cut.

sample	f_b [%]	f_c [%]	f_{light} [%]
<i>b-enriched</i>	15.36 \pm 0.49	12.75 \pm 0.85	71.89 \pm 0.49
<i>c-enriched</i>	4.79 \pm 0.26	10.58 \pm 0.51	84.63 \pm 0.33
<i>light-enriched</i>	1.505 \pm 0.021	4.032 \pm 0.049	94.463 \pm 0.035

Table 4.1: Flavour composition of the calibration samples prepared for the $Z + jet$ analysis. Fractions are measured at the end of the calibration procedure, and reported with their statistical uncertainty.

The di-jet events used to build the calibration samples are selected among the data events acquired with jet triggers. In particular the JET20 and JET50 trigger path are exploited, which are defined to acquire events with at least one jet with transverse energy E_T greater than 20 or 50 GeV respectively. The combination of these two data samples provides a suitable statistics for the calibration procedure. The JET50 sample is particularly useful to enrich the calibration samples of high-energy jets. The E_T distributions of the most energetic jet for JET20 and JET50 data samples are shown in figure 4.1.

Di-jet events are selected in the JET20 data sample according to the following requirements:

- *probe-jet* j_p : $E_T > 20$ GeV and $|\eta| < 1.5$;
- *away-jet* j_a : $E_T > 30$ GeV and $|\eta| < 1.5$;
- $\Delta\phi(j_p, j_a) > 2$;

The cut on the *away-jet* E_T is higher of the trigger threshold in order to remove any bias due to the trigger selection on the *probe-jet* distributions.

Di-jet events are selected in the JET50 data sample according to the following requirements:

- *probe-jet* j_p : $E_T > 70$ GeV and $|\eta| < 1.5$;
- *away-jet* j_a : $E_T > 35$ GeV and $|\eta| < 1.5$;
- $\Delta\phi(j_p, j_a) > 2$;

Also in this sample a E_T cut higher than that applied at trigger level is used to remove any trigger bias.

The same selection is applied to both data and MC samples. The selected di-jet samples extracted from JET20 and JET50 data are then added together and divided in the calibration samples. Finally the calibration samples are re-weighted in order to match the *probe-jet* E_T distribution with the measurement sample. In particular, given that the B sample is exploited to extract the *b correction function* (described in section 4.3), this calibration sample is re-weighted to the E_T distribution of the *b* jets in the measurement sample. The same procedure is applied to the C and L calibration samples.

The specific re-weight used for the present analysis is described in section 6.1.3.

4.3 Extraction of the correction functions

The determination of the *correction functions* for the MC templates has been implemented by an iterative procedure which adapts the MC prediction to the calibration samples. The flow diagram of this procedure is sketched in figure 4.3.

In the present analysis the shape of the *correction functions* is chosen to be linear

$$\phi_{jk} = p_{0,j} + p_{1,j} \cdot x_k \quad (4.1)$$

even if the described procedure can be applied to a wider range of functions. In equation 4.1, x_k is the center of the k -th bin of the j -th Neural Network template, and ϕ_{jk} is the corresponding value of the *correction function*.

At the first step all the correction functions are identically 1, i.e. the corrected templates are equal to the original (uncorrected) ones. In each iteration the composition of the B sample is determined by a Likelihood fit in which the measured parameters are the flavour fractions (*b*, *c*, *light*). These fractions are set and a second Likelihood fit of the same calibration sample evaluates the correction function for the *b-template*. The new corrected template is then obtained applying this correction function to the original template. The details of both the fit procedures are discussed in Appendix A. The same procedure is applied to the C and L samples to evaluate respectively the correction func-

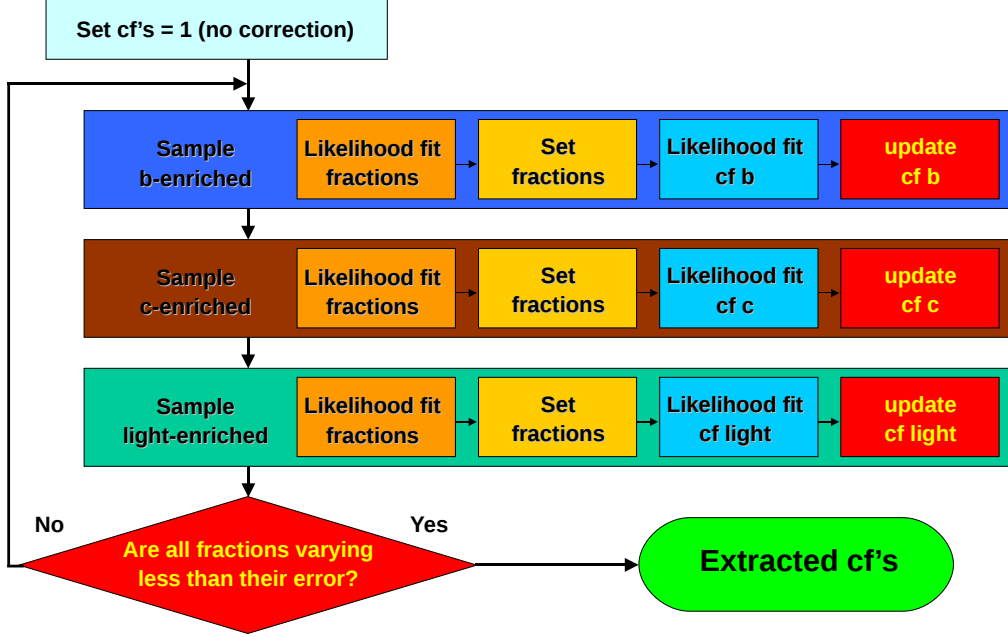


Figure 4.3: Flow diagram of the iterative procedure used to extract correction functions.

tions for *c* and *light* template. Each correction function is evaluated on the calibration sample most enriched in the same flavour to exploit as much as possible the available statistics.

If, at a certain iteration, an evaluated correction function gets negative values inside the range where the tagger output is defined, then it is rejected and replaced by the one obtained on the same calibration sample in the previous iteration.

At the end of each iteration (starting from the second) the variation of the measured fractions respect to their statistical uncertainty is evaluated according to the following *exit criteria* is evaluated:

$$(f_{i,k}^j - f_{i,k}^{j-1}) < \sigma(f_{i,k}^j) \quad \forall i, k \quad (4.2)$$

where $f_{i,k}^j$ is the fraction of the *i* component of the *k* calibration sample, evaluated at the *j*-th iteration, and $\sigma(f_{i,k}^j)$ is its statistical uncertainty. When this condition occurs for all the fractions of all the calibration sample, the iteration

procedure is terminated.

The evolution of the fractions measured in the calibration samples as a function of the iteration is shown in figure 4.4 for the $Z + jet$ analysis (calibration samples reweighted to the E_T distribution of $Z + jet$ events). The corresponding evolution of the component fits of the calibration samples is reported in figures 4.5, 4.6 and 4.7.

The effect of the calibration procedure on b , c and *light* templates is shown in figure 6.4 for the $Z + jet$ analysis.

Several tests have been performed to evaluate the stability of the developed iterative procedure:

- changing the order of the enriched samples inside the iterative procedure, but always extracting the b cf from B sample, and so forth, does not affect the final cf set.
- The procedure is also checked to converges to the same results in terms of cf , regardless the input fractions used at the first step. In fact the iterative procedure converges to the same cf set even if the calibration samples fractions measured at the first step are substituted with the corresponding MC prediction.

The stability of the iterative procedure, and the possibility of tuning the calibration samples allow to exploit the presented data-driven calibration technique for a wide range of high- p_T analysis. The procedure developed to evaluate the indeterminations associated to this technique is presented in section 4.4. The specific application to the $Z + jet$ analysis is presented in chapter 6.

4.4 Uncertainty of the correction functions

In section 4.3 the template calibration procedure developed to the determine a set of three *correction functions* for pure b , c and *light* Neural Network tagger distributions, has been presented. Given a specific set of enriched data and Monte Carlo samples, the uncertainty on correction functions parameters

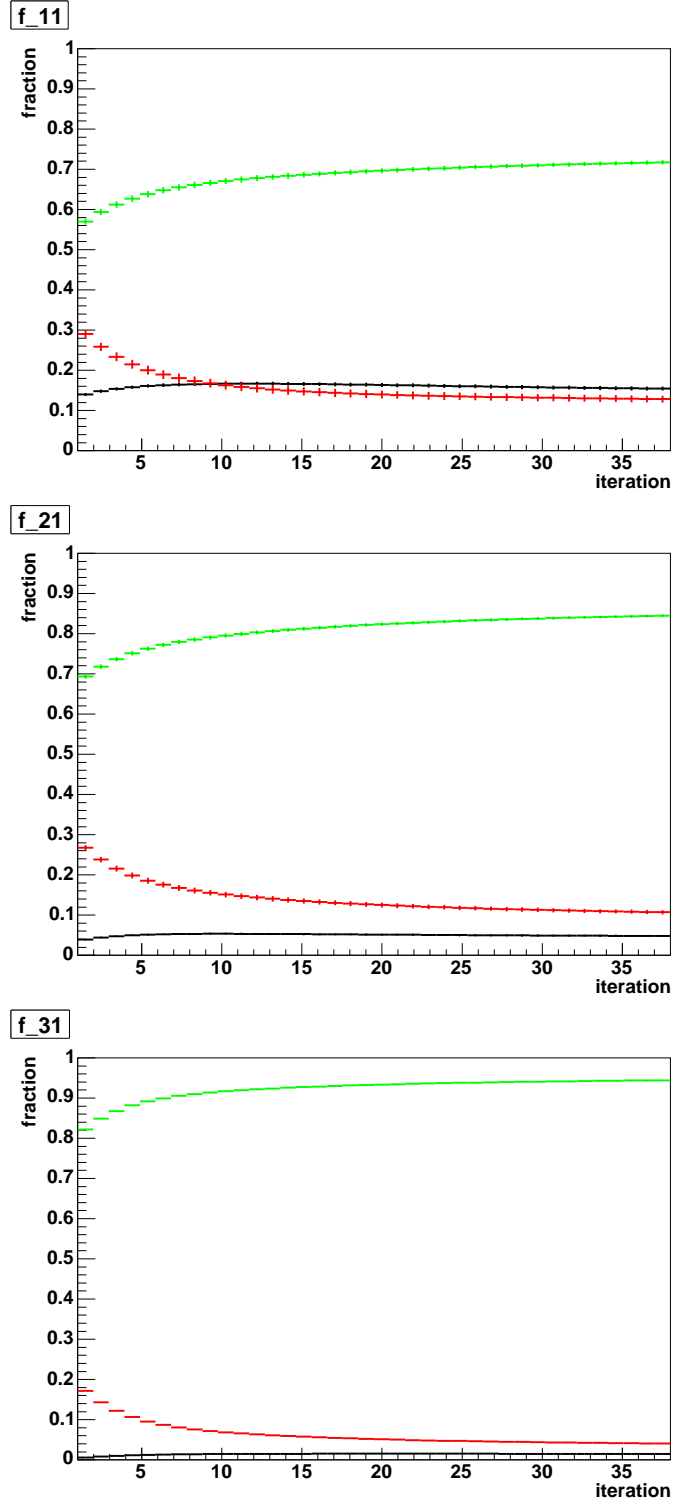


Figure 4.4: Evolution of the b (black), c (red) and $light$ (green) fractions measured in the calibration samples as a function of the iteration in the b -enriched (upper panel), c -enriched (central panel) and $light$ -enriched (lower panel) samples.

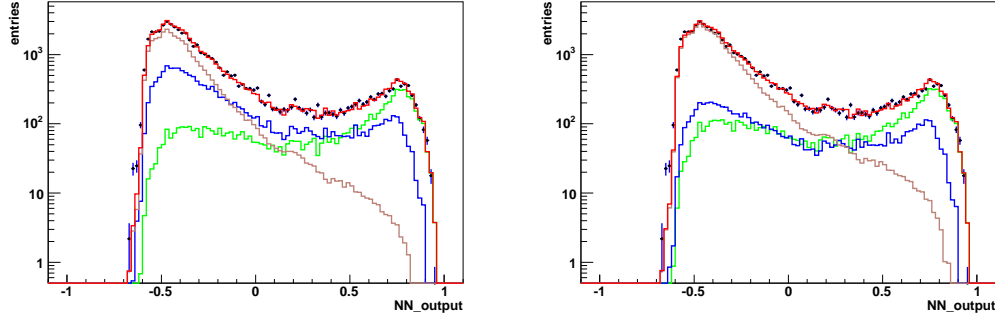


Figure 4.5: Neural Network distribution of the *b-enriched* sample compared with the corresponding component fit for the first (left panel) and the last (right panel) iteration of the calibration procedure.

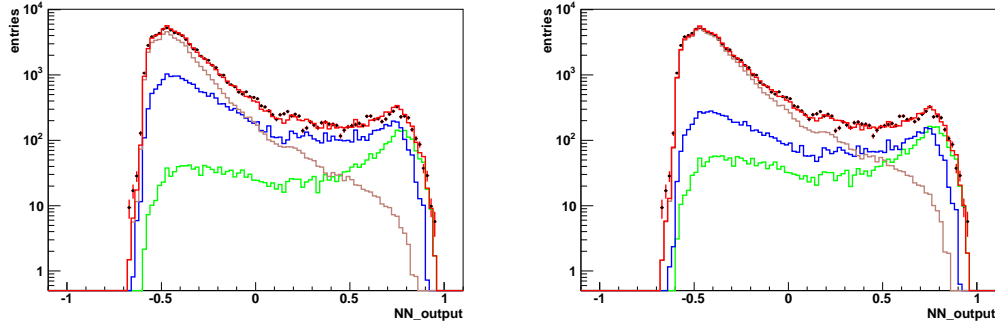


Figure 4.6: Neural Network distribution of the *c-enriched* sample compared with the corresponding component fit for the first (left panel) and the last (right panel) iteration of the calibration procedure.

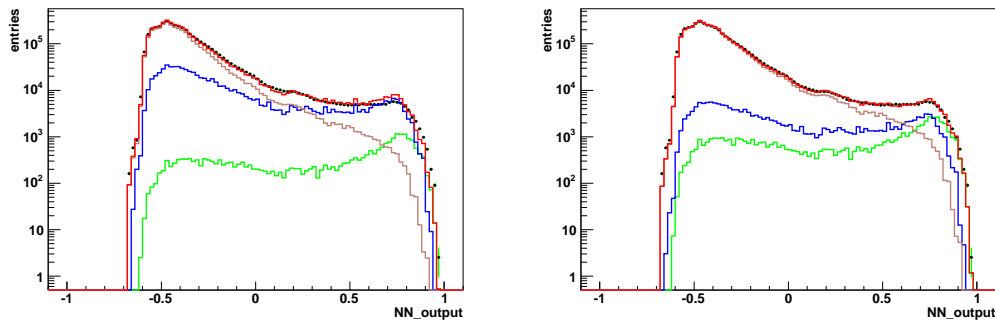


Figure 4.7: Neural Network distribution of the *light-enriched* sample compared with the corresponding component fit for the first (left panel) and the last (right panel) iteration of the calibration procedure.

ultimately depends on the features of the used NN distributions in terms of statistical fluctuations of the bins content, for both data and Monte Carlo templates.

A realistic toy Monte Carlo, reproducing the complete calibration procedure, has been developed in order to evaluate the spread of the cf 's obtained from several pseudo-calibration experiments. A set of $3n$ pseudo-data samples reproducing the three flavour-enriched samples used in the calibration procedure, are extracted from the MC templates according to the original statistics. The b , c , and *light* components of each pseudo-calibration sample are generated according to the corresponding MC template, corrected with its nominal cf and re-weighted to the E_T spectra used in the nominal calibration procedure for to re-weight that enriched sample.

For each of these n sets of pseudo-calibration experiments a complete calibration procedure is performed, producing n sets of three cf 's. The cf 's triplets, generated in this way, contain by construction all the correlations between their parameters coming from the calibration procedure.

These sets of cf 's can be exploited to evaluate the uncertainty on a fractions measurement due to the calibration procedure. To this purpose a fractions fit can be repeated n times on the measurement sample, using every time a different set of cf 's: the spread of the measured fractions quantifies the uncertainty on the fractions due to the indetermination of the cf 's.

Chapter 5

Z+jets events selection

Events of associated production of a Z boson and at least one jet are identified in the data sample acquired by the CDF detector. The Z boson identification is performed reconstruction its leptonic decay in e^+e^- and $\mu^+\mu^-$ pairs. The jets are reconstructed using the JETCLU cone algorithm. Background events to the $Z + jet$ are evaluated exploiting both data and MC samples.

The selected sample is used to perform the Heavy Flavour fraction measurement described in section 6.

5.1 Data and Monte Carlo samples

Data

The data samples analyzed in the present work have been collected exploiting the high- P_T electrons and muons trigger paths. The available integrated luminosity is 2.05 fb^{-1} for the electron trigger and 1.90 fb^{-1} for the muon trigger, collected between February 2002 and May 2007 (run number range: 138425 - 241664).

The ELECTRON_CENTRAL18 trigger path, which requests an electron with $E_T > 18 \text{ GeV}$, has been used for the research of $Z \rightarrow e^+e^-$ events. The cuts applied by this trigger path are reported in table 5.1. For the research of $Z \rightarrow \mu^+\mu^-$ events, a set of high- p_T muon trigger paths, which request a muon with $P_T > 18$

GeV/ c in the central muon detectors, has been used. The cuts applied by the two main high- p_T muon trigger paths used (MUON_CMUP18 and MUON_CMX18) are reported in table 5.2 and 5.4 respectively.

Monte Carlo

The signal Monte Carlo samples have been generated using Alpgen v2.10, for the ME calculation, and Pythia v6.325 for the PS simulation. The sample generated are divided according to the parton multiplicity (0, 1, 2, 3 or ≥ 4 partons) and flavour: $Z+b\bar{b}+np$, $Z+c\bar{c}+np$ and $Z+light\ p$. The signal events are generated in 75-105 GeV/ c^2 boson invariant mass range. The division in flavour sub-sample allows for a direct double counting removal of the events with Heavy Flavour quarks generated by Pythia in the PS simulation, rather than by Alpgen in the ME calculation. To this purpose, the following procedure is implemented:

- events with a jet containing a B or C hadron are removed from the $Z+light\ p$ samples;
- events with a jet containing a B hadron are removed from the $Z+c\bar{c}+Xp$ samples.

Further events are generated for Drell-Yan processes Z/γ^*+np in the 20-75 GeV/ c^2 and 105-600 GeV/ c^2 ranges for the boson invariant mass (*sidebands*).

Given that Alpgen generator performs the ME calculation at leading order (LO), the cross sections evaluated for the subsamples need to be corrected to match the next-to-leading order (NLO) calculation or the experimental measurement. The ratio between the NLO and LO cross sections for a given process is the so-called k -factor:

$$k = \frac{\sigma_{NLO}}{\sigma_{LO}} \quad (5.1)$$

The k -factor is evaluated to be 1.37 for Alpgen+Pythia simulations, and 1.4 when Pythia is used for the whole simulation. All the MC subsamples are then scaled to the data integrated luminosity \mathcal{L} :

$$w_i = \frac{k \cdot \sigma_{Alpgen}}{N_i^g} \cdot \mathcal{L} \quad (5.2)$$

Trigger Name	Requirements
L1_CEM8_PT8	≥ 1 trigger tower $ET_CENTRAL \geq 8 \text{ GeV}$ $HAD_EM_CENTRAL \leq 0.125$ (applied only if $ET_CENTRAL < 14 \text{ GeV}$) $XFT_PT \geq 8.34 \text{ GeV}/c$ $XFT_LAYERS \geq 4$ $XFT_CHARGE = 0$
L2_CEM16_PT18	≥ 1 trigger tower $L2_EM_T \geq 16 \text{ GeV}$ $L2_HAD_EM_RATIO \leq 0.125$ $DCAS_HIGH_EM_CENTRAL_SEED \geq 8 \text{ GeV}$ $DCAS_HIGH_EM_CENTRAL_SHOULDER \geq 7.5 \text{ GeV}$ $XFT_PT \geq 8 \text{ GeV}/c$ $ABS_ETA_MAX = 1.317$
L3_ELECTRON_CENTRAL18	$nEmObj \geq 1$ $CalorRegion = 0$ $CenET > 18 \text{ GeV}$ $cenHadEm \leq 0.125$ $Ishr \leq 0.4$ $cenDeltaZ \leq 8 \text{ cm}$ $ZVert = 2$ $nTowersHadEm = 3$ $cenTrackPt \geq 9.0 \text{ GeV}/c$

Table 5.1: Description of the ELECTRON_CENTRAL18 trigger path

MUON_CMUP18 trigger path	
Level 1	L1_CMUP6_PT4
Level 2	L2_AUTO_L1_CMUP6_PT4 ($run \leq 152949$) L2_TRK8_L1_CMUP6_PT4 ($152949 < run \leq 179056$) L2_CMUP6_PT8 ($181013 \leq run \leq 202717$) L2_CMUP6_PT15 ($run \geq 198428$)
Level 3	L3_MUON_CMUP18

Table 5.2: Description of the MUON_CMUP18 trigger path.

Trigger Name	Requirements
L1_EM8_&_MET15	≥ 1 trigger tower $ET_CENTRAL$ (or PLUG) ≥ 8 GeV $HAD_EM_CENTRAL$ (or PLUG) ≤ 0.125 $MET \geq 15$ GeV $SUMET_THRESH_CENTRAL$ (or PLUG) ≥ 1 GeV
L2_CEM16_L1_MET15	$L2_EM_ET \geq 16$ (20) GeV $L2_HAD_EM_RATIO \leq 0.125$ $DCAS_HIGH_EM_CENTRAL_SEED \geq 8$ GeV $DCAS_HIGH_EM_CENTRAL_SHOULDER \geq 7.5$ GeV $ABS_ETA_MAX = 1.1$ $L1_MET \geq 15$ GeV
L3_W_NOTRACK_MET25	$nEmObj \geq 1$ $CalorRegion = 2$ $CenET$ (or plug) > 25 GeV $cenHadEm$ (or plug) ≤ 0.125 $nTowersHadEm = 3$ $MET \geq 25$ GeV

Table 5.3: Path di trigger per il W_NOTRACK

MUON_CMX18 trigger path	
Level 1	$L1_CMX_PT8_PS1$ (run ≤ 152949) $L1_CMX_PT8_CSX(_{PS1})$ (run > 152949)
Level 2	$L2_AUTO_L1_CMX6_PT8$ (run ≤ 152949) $L2_AUTO_L1_CMX6_PT8_CSX$ ($152525 \leq \text{run} \leq 179056$) $L2_AUTO_L1_CMX6_PT10$ ($181013 \leq \text{run} \leq 202717$) $L2_L1_CMX6_PT15$ (run ≥ 198428)
Level 3	$L3_MUON_CMX18$

Table 5.4: Description of the MUON_CMX18 trigger path.

where N_i^g is the number of events generated for the i subsample (after the double counting removal) and w_i is the weight of the i subsample.

Samples of di-boson events in associated production with partons, ($ZZ+np$, $ZW+np$, $WW+np$) are simulated using Alpgen+Pythia combination.

Samples of $t\bar{t}$ events (with $m_t = 175 \text{ GeV}/c^2$) are simulated using Pythia, and are normalized to the value of the predicted value [28] of the production cross section $\sigma(p\bar{p} \rightarrow t\bar{t}) = 6.7 \text{ pb}$.

The selection described in the next sections is applied to both data and MC samples. The selected MC events are weighted according to the trigger efficiency, because the trigger simulation is not implemented for MC samples.

Z boson selection

The identification of the Z boson relies on the reconstruction of its leptonic decay in e^+e^- and $\mu^+\mu^-$ pairs. In the next sections the procedures for high- P_T electrons and muons identification and the corrections to the Monte Carlo simulations for trigger and selection efficiencies, are discussed. The result of the Z boson selection are presented with the corresponding Monte Carlo prediction.

5.2 Electron selection

The identification of an electron candidate is based on the reconstruction of a cluster in the electromagnetic calorimeter. Further information can be added by a track matching the electromagnetic cluster, reconstructed by the central tracker (COT) in the *central* region ($|\eta| < 1$), and by the preshower system (PPR) and the shower max detector (PES) in the *plug* region ($1 < |\eta| < 3.6$). The reconstructed quantities used to select the electrons are briefly described in the following; the corresponding distributions are shown in figures from 5.1 to 5.6.

Variable	Tight Central	Loose Central	Plug
Detector	CEM	CEM	PEM
Fiducial	Fid=1	Fid=1	-
E_T	$> 18 \text{ GeV}$	$> 18 \text{ GeV}$	$> 18 \text{ GeV}$
Conversion	No conversion	-	-
Track $ z_0 $	$\leq 60 \text{ cm}$	-	-
TrkAxSeg	$\geq 3 \text{ SL}$ with 5 hits SL	-	-
TrkStSeg	$\geq 2 \text{ SL}$ with 5 hits SL	-	-
Track P_T	$\geq 9 \text{ GeV/c}$	$\geq 9 \text{ GeV/c}$	-
Had/Em	$\leq 0.055 + 0.00045E$	$\leq 0.055 + 0.00045E$	$\leq 0.055 + 0.00045E$
Iso E_T/E_T	≤ 0.1	≤ 0.1	≤ 0.1
Lshr	≤ 0.2	-	-
E/P	≤ 2 unless $P_T > 50 \text{ GeV/c}$	-	-
CES $ \Delta z $	$\leq 3.0 \text{ cm}$	-	-
CES $q \cdot \Delta x$	$-3 \leq q \cdot \Delta x \leq 1.5$	-	-
CES χ_{strip}^2	≤ 10	-	-

Table 5.5: Summary of the electron selection cuts.

- Calorimetric variables
 - E_T : projection on the transverse plane ($r - \phi$) of the energy measured by the calorimeter;
 - Had/Em : ratio of the energy deposited by the candidate electron in the hadronic (Had) and in the electromagnetic (Em) calorimeters (figure 5.1). The cut applied on this variable is parametrized as a function of the electromagnetic energy E_{Em} in order to reduce the dependence of its efficiency on the electron energy.
 - *Isolation* : it is a measurement of the calorimetric activity around the candidate electron (figure 5.1). For the electron identification, the isolation is defined as:

$$Isol = \frac{E_T^{0.4} - E_T^{cluster}}{E_T^{cluster}} \quad (5.3)$$

where $E_T^{0.4}$ is the sum of the transverse energy of the towers inside a cone with a radius $R = 0.4$ around the cluster axis, and $E_T^{cluster}$ is the energy of the cluster associated to the candidate electron.

- $Lshr$: this variable quantifies a comparison between the measured cluster profile with the one expected from test-beam data:

$$Lshr = 0.14 \cdot \sum_j \frac{E_j - E_j^{exp}}{\sqrt{(0.14\sqrt{E_i})^2 + (\Delta E_j^{exp})^2}} \quad (5.4)$$

where E_j is the measured energy in the j tower, E_j^{exp} is the expected value for the energy in the j tower, according to a parametrization based on test-beam data, and the index j runs over the tower close to the *seed tower* (figure 5.2).

- χ_{strip}^2 : it is the χ^2 of the fit to the energies of the 11 CES strips associated to a cluster, according to a parametrization extracted from test-beam data (figure 5.4).

- Track variables

- $TrkAxSeg$ and $TrkStSeg$: these are the numbers of axial and stereo super-layers, respectively, with at least 5 hits (figure 5.3).
- E/p : ratio of the energy of the calorimetric cluster with the momentum measured by the central tracker.
- Δz : separation in the $z - \phi$ plane between the position of the cluster measured by the CES and the extrapolation of the track to the CES plane (figure 5.4).
- $q \cdot \Delta x$: separation in the $r - \phi$ plane between the position of the cluster measured by the CES and the extrapolation of the track to the CES plane, multiplied by the charge of the candidate electron q (figure 5.5). The $r - \phi$ plane is the bending plane of the magnetic field produced by the solenoid, so this variable presents an asymmetric distribution.

In the present analysis three categories of identified electrons are used: *tight* and *loose* in the central region, and *plug* in the forward region.

The cuts applied for each category of electrons are reported in table 5.5.

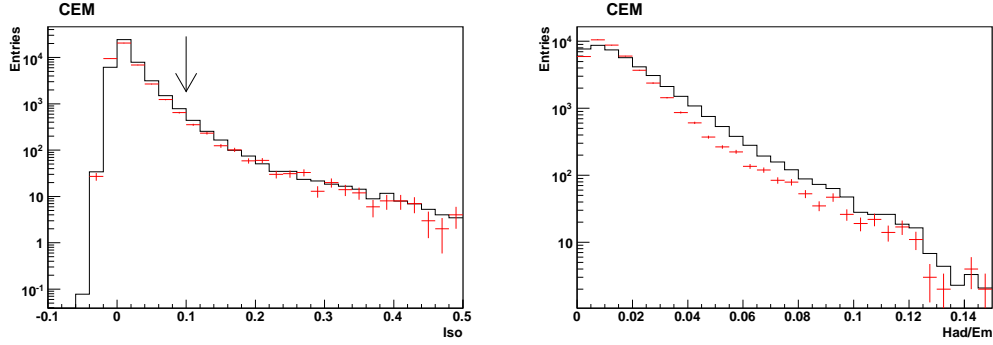


Figure 5.1: Comparison between data (red) and MC (black) of $Isolation$ (left panel) and Had/Em (right panel) distributions for central electrons. The arrow indicates the standard identification cut.

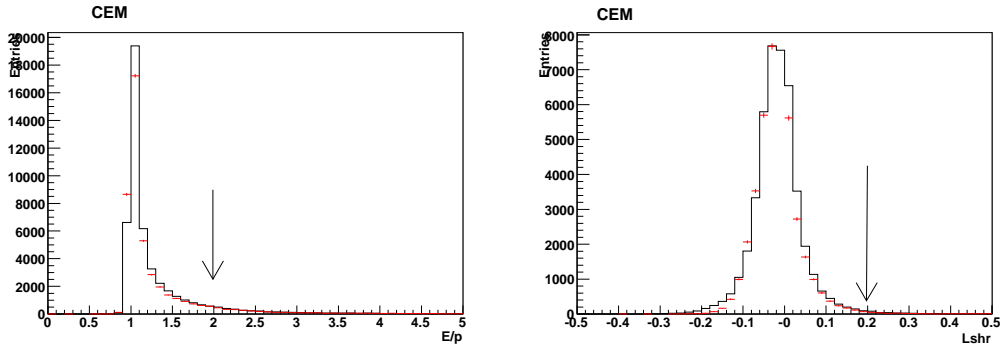


Figure 5.2: Comparison between data (red) and MC (black) of E/p (left panel) and $Lshr$ (right panel) distributions for central electrons. The arrows indicate the standard identification cut.

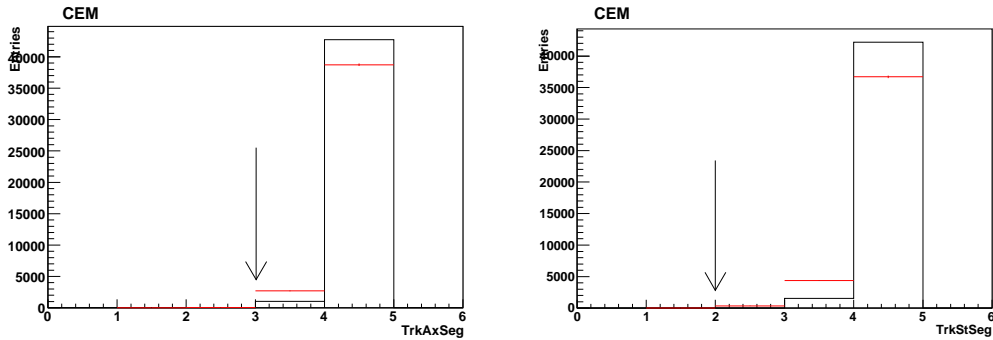


Figure 5.3: Comparison between data (red) and MC (black) of the number of axial (left panel) and stereo (right panel) COT segments distributions for central electrons. The arrows indicate the standard identification cut.

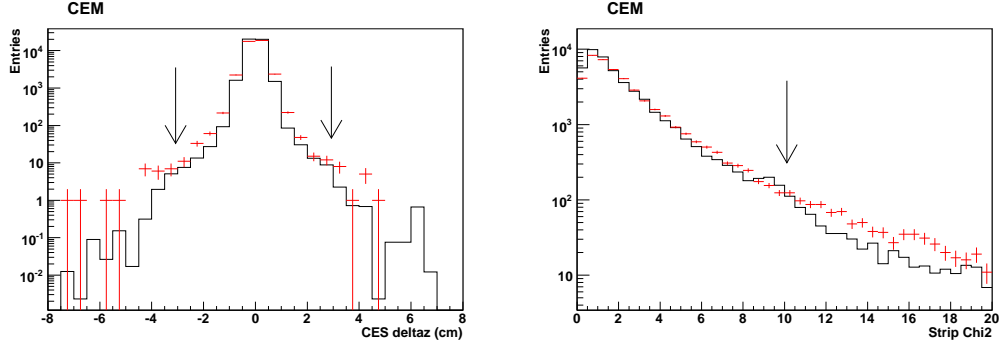


Figure 5.4: Comparison between data (red) and MC (black) of Δz (left panel) and χ^2_{strip} (right panel) distributions for central electrons. The arrows indicate the standard identification cut.

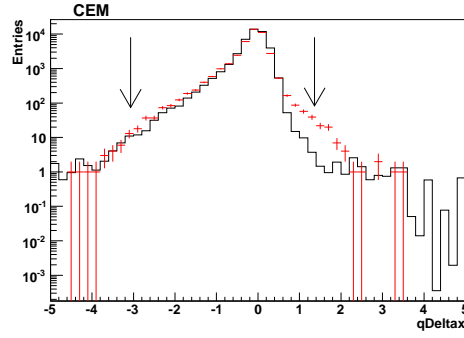


Figure 5.5: Comparison between data (red) and MC (black) of $q\Delta x$ distribution for central electrons. The arrows indicate the standard identification cut.

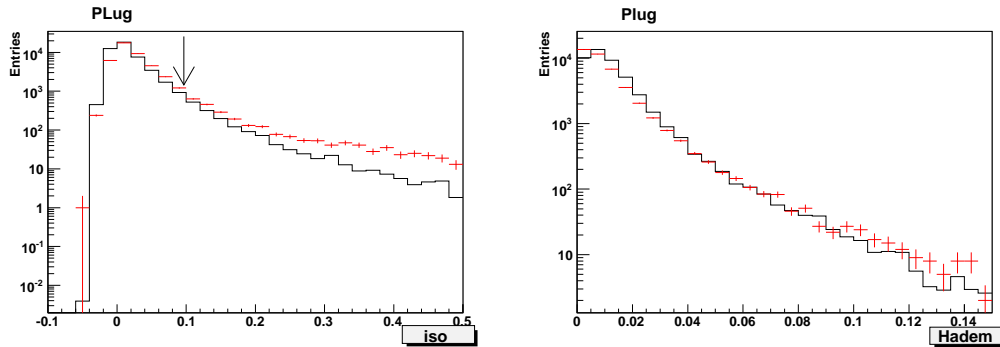


Figure 5.6: Comparison between data (red) and MC (black) of *Isolation* (left panel) and *Had/Em* (right panel) distributions for plug electrons. The arrows indicate the standard identification cut.

Run range	$\epsilon_{Tracking}$	$\epsilon_{Calorimeter}$
138425 - 212133	0.979	$1 - 59106 \cdot e^{-0.70 \cdot E_T}$
217990 - 222426	0.958	$1 - 85036 \cdot e^{-0.71 \cdot E_T}$
222529 - 228596	0.9593	$\frac{1}{1+e^{-0.5 \cdot E_T - 13.0}} \cdot \frac{1}{1+e^{-0.27 \cdot E_T}}$
228664 - 241664	0.9596	$\frac{1}{1+e^{-1.38 \cdot E_T - 17.88}}$

Table 5.6: Summary of the tracking ($\epsilon_{Tracking}$) and calorimetric ($\epsilon_{Calorimeter}$) efficiencies for the ELECTRON_CENTRAL18 trigger path.

5.2.1 Electron trigger efficiency

The standard method adopted by CDF to measure the trigger efficiency exploits an *unbiased* data sample, acquired with an independent trigger. The trigger ELECTRON_CENTRAL18, used for the present analysis, exploits both tracking and calorimetric information, and the corresponding contributions to the trigger efficiency can be evaluated separately.

The tracking efficiency can be evaluated in a data sample acquired with a trigger path which implements the same calorimeter requests of the ELECTRON_CENTRAL18, and has no requests on tracking quantities. A suitable trigger for this study is W_NOTRACK, described in table 5.3. The tracking efficiency varies with the η of the calorimetric cluster, as shown in figure 5.7. This dependence is ignored in the present analysis, and the resulting systematic uncertainty ($\sim 1.5\%$) on signal MC normalization is included in the uncertainty on the integrated luminosity measurement (6%).

The calorimetric efficiency can be evaluated in the *tight* electrons sample acquired with the ELECTRON_CENTRAL18 trigger. Due to the structure of the tower clustering algorithm implemented in the level 2 of this trigger path, the calorimetric efficiency is a function of the electron transverse energy, as shown in figure 5.8, and it is parametrized with an exponential function.

The values of the tracking and calorimetric efficiencies of the ELECTRON_CENTRAL18 trigger path used in the present analysis are reported in table 5.6, according to the application run range. The total trigger efficiency for a central event is

$$\epsilon_{trig}^e = \epsilon_{Tracking} \cdot \epsilon_{Calorimeter} \quad (5.5)$$

Run range	<i>Tight</i>	<i>Loose (non-Tight)</i>	<i>Plug</i>
138425 - 186598	0.991	1.001	0.939
190697 - 203799	0.985	1.003	0.941
203819 - 222426	0.974	0.996	0.924
222529 - 228596	0.978	0.996	0.919
228664 - 241664	0.978	0.996	0.915

Table 5.7: Identification efficiency scale factors for different run ranges and non-overlapping electron categories. The slightly decreasing trend is related to the increasing instantaneous luminosity delivered by Tevatron.

5.2.2 Electron selection efficiency

The procedure used by CDF collaboration to evaluate the selection efficiency of the different electron categories is based on a very pure $Z \rightarrow e^+e^-$ sample. These events are identified through the reconstruction of a pair of candidate electrons with invariant mass in a narrow window around the Z mass (76-106 GeV/ c^2), with the first one satisfying very tight identification cuts. The second electron is then exploited for the evaluation of the selection efficiencies for the different set of cuts.

Any disagreement between data and Monte Carlo simulation can reflect in a different value of the selection efficiencies. The standard correction procedure adopted by the CDF collaboration relies on the evaluation of *scale factors* to reconcile the selection efficiencies measured in MC simulation with the ones measured in data samples:

$$SF = \frac{\epsilon_{sel}^{Data}}{\epsilon_{sel}^{MC}} \quad (5.6)$$

The *scale factors* are evaluated for non-overlapping categories, therefore the *tight* electrons are removed from the *loose* sample. The *scale factors* applied in the present analysis to correct the MC selection are reported in table 5.7.

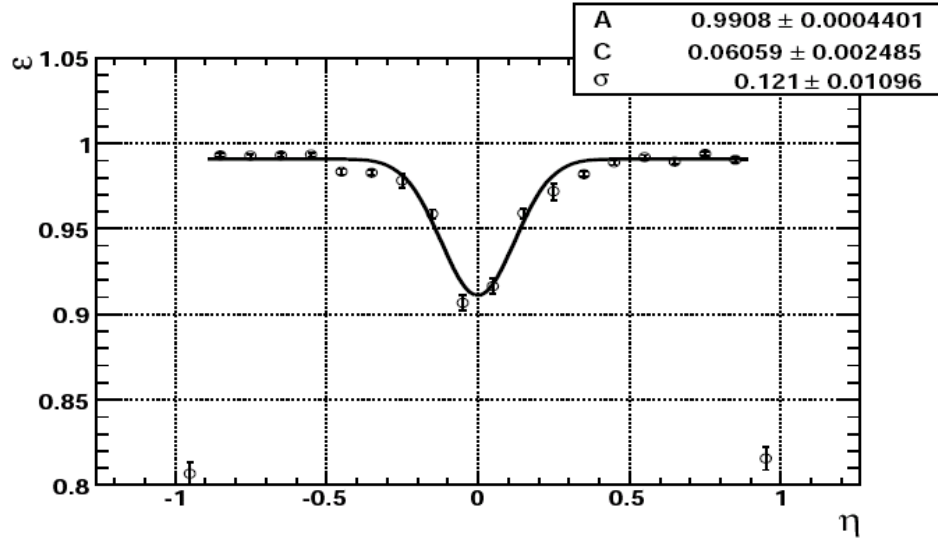


Figure 5.7: Tracking efficiency for ELECTRON_CENTRAL18 trigger as a function of the η of the calorimetric cluster.

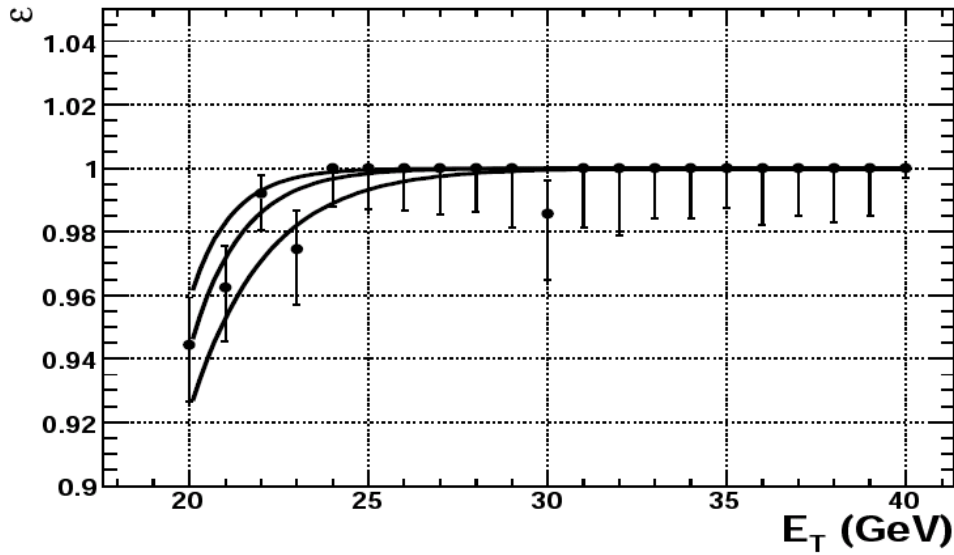


Figure 5.8: Level 2 calorimetric efficiency for ELECTRON_CENTRAL18 trigger path, as a function of the electron transverse energy. The measured values of the efficiency are reported with the statistical error. The nominal fit is superimposed to the distribution, together with the systematic variation range

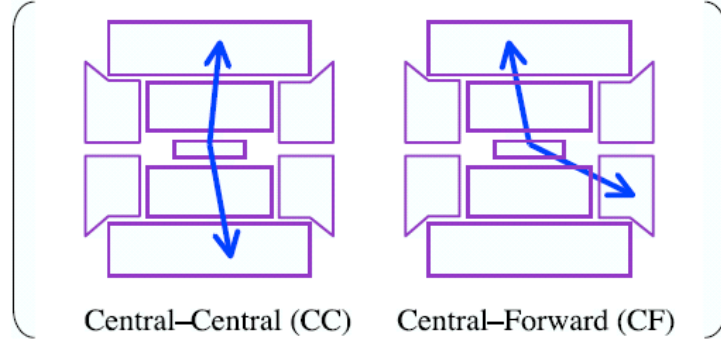


Figure 5.9: Schematic draw of events with two central electrons and with one central and one forward electron.

5.2.3 $Z \rightarrow e^+e^-$ events selection

Events with a Z boson decaying in an electron-positron pair are identified according to the following requests:

- a first $e^{+/-}$ satisfying *tight* cuts;
- a second $e^{+/-}$ satisfying *loose* or *plug* cuts;
- the invariant mass of the lepton pair in the 66-116 GeV/c^2 range.

No explicit opposite-charge request is applied to the lepton pair given the low electron charge identification efficiency in the central region of the detector, and to the absence of the reconstructed track in the forward region.

According to the set of cuts satisfied by the second lepton, the selected events can be divided in *central-central* (CC) and *central-plug* (or *central-forward*) (CP), as schematically shown in figure 5.9.

The same selection is applied to both data and MC samples. Given that the MC simulation does not include the trigger simulation, each selected MC event is weighted according to the combined trigger efficiency of the two leptons. In CC events, both the electrons can fire the trigger, so the combined trigger

efficiency ϵ_{CC}^{trig} is given by

$$\epsilon_{CC}^{trig} = 1 - (1 - \epsilon_1^{trig}) \cdot (1 - \epsilon_2^{trig}) \quad (5.7)$$

where ϵ_1^{trig} and ϵ_2^{trig} are the trigger efficiencies for the selected central electrons, evaluated event by event according to the scheme in section 5.2.1. On the other hand, in CP events only the *tight* electron hits the detector in the central region, therefore the trigger efficiency ϵ_{CP}^{trig} is given by

$$\epsilon_{CP}^{trig} = \epsilon_{tight}^{trig} \quad (5.8)$$

Also the selection efficiency of the $Z \rightarrow e^+e^-$ events can be evaluated combining the selection efficiencies of the selected leptons. The CC events can be divided in two categories according to the set of cut satisfied by the second lepton: *tight* or *loose-non-tight*. Combining these two categories, the selection efficiency for CC events ϵ_{CC}^{sel} results

$$\epsilon_{CC}^{sel} = \epsilon_{tight}^{sel} \cdot \epsilon_{tight}^{sel} + 2\epsilon_{loose}^{sel} \cdot (\epsilon_{loose}^{sel} - \epsilon_{tight}^{sel}) \quad (5.9)$$

The combined selection efficiency for CP events is given by

$$\epsilon_{CP}^{sel} = \epsilon_{tight}^{sel} \cdot \epsilon_{plug}^{sel} \quad (5.10)$$

The selection efficiency *scale factors* for CC and CP events respectively are given by:

$$SF_{CC} = SF_{tight} \cdot SF_{loose-non-tight} \quad (5.11)$$

$$SF_{CP} = SF_{tight} \cdot SF_{plug} \quad (5.12)$$

5.2.4 Electron energy scale

Any miscalibration of the electron absolute energy scale modifies the invariant mass of the selected lepton pair. The measured invariant mass of the candidate Z boson is corrected both in data and MC simulation according to the procedure described in the following. Considering the four-momentum of

the selected leptons:

$$\vec{e}_1 = (E_1, p_1^x, p_1^y, p_1^z) \quad (5.13)$$

$$\vec{e}_2 = (E_2, p_2^x, p_2^y, p_2^z) \quad (5.14)$$

the invariant mass of the lepton pair can be expressed as

$$M_{ee} = \sqrt{(E_1 + E_2)^2 - (p_1^x + p_2^x)^2 - (p_1^y + p_2^y)^2 - (p_1^z + p_2^z)^2} \quad (5.15)$$

$$= \sqrt{2E_1E_2(1 - (\cos\phi_1 \cdot \cos\phi_2 + \sin\phi_1 \cdot \sin\phi_2) \cdot \sin\theta_1 \cdot \sin\theta_2 - \cos\theta_1 \cdot \cos\theta_2)} \quad (5.16)$$

$$= \sqrt{2E_1E_2 \cdot f(\phi_1, \theta_1, \phi_2, \theta_2)} \quad (5.17)$$

where ϕ and θ are the polar and azimuthal angle respectively in the CDF detector reference frame. The reconstructed invariant mass can be scaled to the Z boson mass through a factor α :

$$M_Z = \alpha \cdot M_{ee} = \sqrt{2 \cdot \alpha^2 \cdot E_1E_2 \cdot f(\phi_1, \theta_1, \phi_2, \theta_2)} \quad (5.18)$$

The α^2 factor can be expressed as a product of two scaling factors ($K_1 \cdot K_2$) for the energies of the two leptons:

$$E'_1 = K_1 \cdot E_1 \quad (5.19)$$

$$E'_2 = K_2 \cdot E_2 \quad (5.20)$$

Assuming a constant values of the energy scale factor for central (K_c) and plug (K_p) electrons, the equation 5.18 becomes for CC events:

$$M_Z = \sqrt{2 \cdot K_c^2 \cdot E_1E_2 \cdot f(\phi_1, \theta_1, \phi_2, \theta_2)} \quad (5.21)$$

and for CP events:

$$M_Z = \sqrt{2 \cdot K_c \cdot K_p \cdot E_1E_2 \cdot f(\phi_1, \theta_1, \phi_2, \theta_2)} \quad (5.22)$$

Substituting 5.17 in 5.21 and 5.22 for CC and CP events respectively:

$$M_Z = K_c \cdot M_{ee}^{CC} \quad (5.23)$$

$$M_Z = \sqrt{K_c \cdot K_p} \cdot M_{ee}^{CP} \quad (5.24)$$

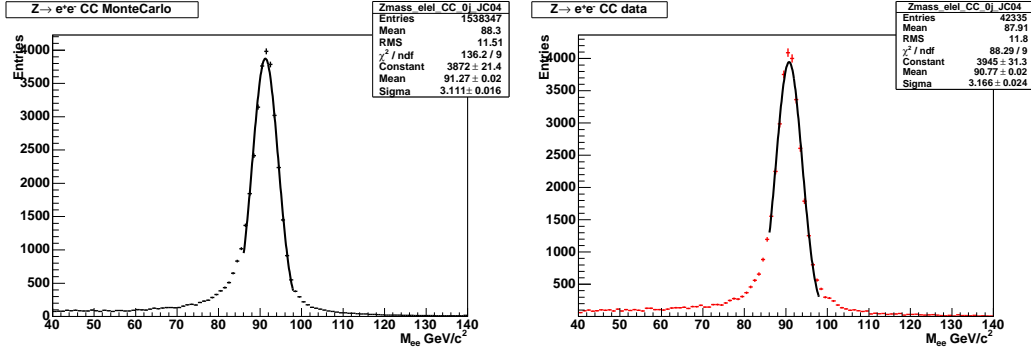


Figure 5.10: Distribution of the e^+e^- invariant mass (M_{ee}) for *central-central* events in data (left panel) and MC (right panel). A gaussian fit in the 86-98 GeV/c^2 range is superimposed.

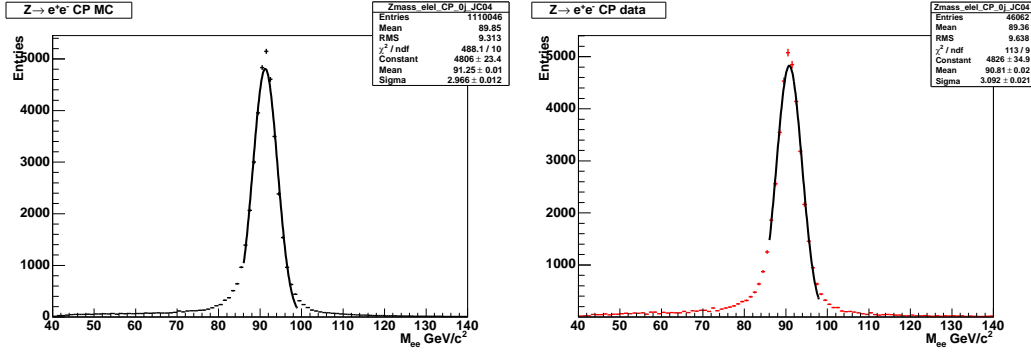


Figure 5.11: Distribution of the e^+e^- invariant mass (M_{ee}) for *central-plug* events in data (left panel) and MC (right panel). A gaussian fit in the 86-98 GeV/c^2 range is superimposed.

η range		MC	Data
-2.80	$< \eta_e < -1.78$	0.998	1.004
-1.78	$< \eta_e < -1.20$	0.996	1.004
-1.00	$< \eta_e < 1.00$	0.996	1.002
1.20	$< \eta_e < 1.78$	0.996	1.004
1.78	$< \eta_e < 2.80$	0.998	1.004

Table 5.8: Electron energy scale factors for data and Monte Carlo simulation, evaluated in different electron η ranges.

Resolving 5.23 and 5.24 for K_c and K_p :

$$K_c = \frac{M_Z}{M_{ee}^{CC}} \quad (5.25)$$

$$K_p = \frac{M_Z \cdot M_{ee}^{CC}}{(M_{ee}^{CP})^2} \quad (5.26)$$

The values of M_{ee}^{CC} and M_{ee}^{CP} are then extracted from a gaussian fit to the distribution of the invariant mass of the selected lepton pair for CC and CP events, in the 86-98 GeV/ c^2 range and for different electron η ranges, as shown in figures 5.10 and 5.11. The electron energy scale factors are evaluated both in data and MC simulation (table 5.8), and applied in the selection procedure.

5.2.5 Background to $Z \rightarrow e^+e^-$ events

The main source of background events for the $Z \rightarrow e^+e^-$ process is represented by QCD or $W + jets$ events in which one or more jets are reconstructed as electrons. Other appreciable sources of background events come from those processes involving a real Z boson or in which two leptons can be produced with invariant mass close to Z mass, like di-boson (ZZ , ZW and WW) and $t\bar{t}$ events.

QCD and $W + jet$

The QCD and $W + jet$ events, with the W decaying in $e\nu$, are the main background to the $Z + jet$ signal. In both cases one or more jets can be misidentified as electrons (*fake electrons*), generating a signature similar to signal.

The procedure used to evaluate this kind of background is data driven. The probability of a jet to be reconstructed as an electron is quantified by the *fake rate*, that is defined as

$$fake\ rate(\lambda) = \frac{\text{number of selected jets according the } \lambda \text{ selection cuts}}{\text{number of jets in the control sample}}$$

where λ is the set of electron selection cuts applied (*tight*, *loose* or *plug*). The control sample is defined as:

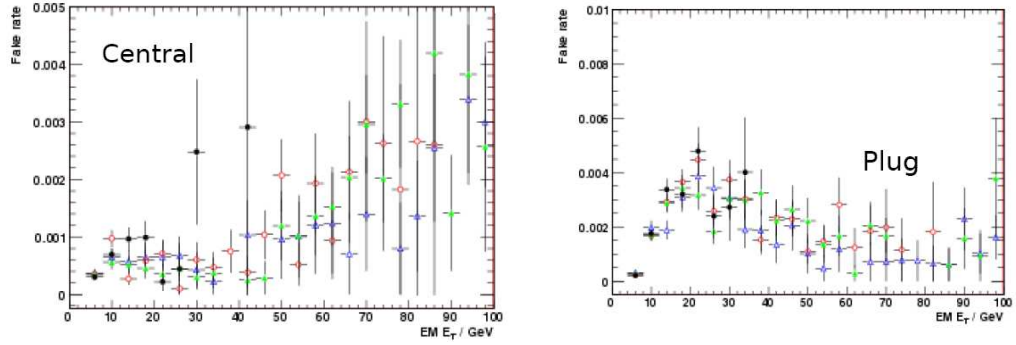


Figure 5.12: *Fake rate* as a function of the jet transverse energy, measured in the inclusive jet samples for the central (left panel) and plug (right panel) jets.

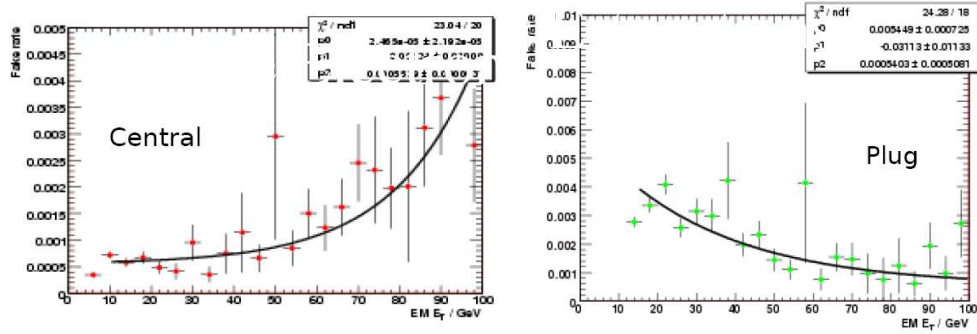


Figure 5.13: Average *fake rate* as a function of the jet transverse energy, with the exponential fit superimposed, for the central (left panel) and plug (right panel) jets.

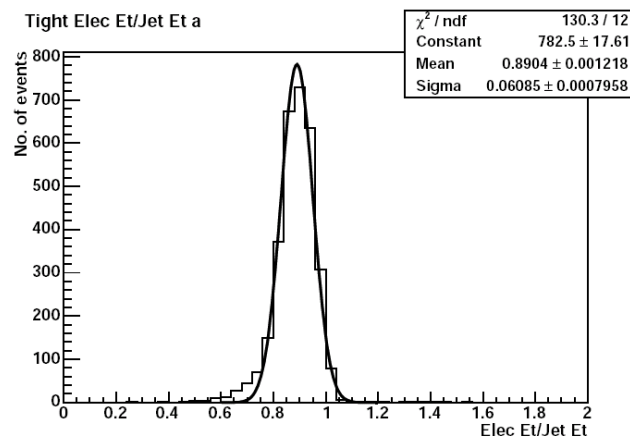


Figure 5.14: Distribution of the ratio of the *fake electron* energy and the corresponding jet energy, with a gaussian fit superimposed.

- the number of central jets ($|\eta| < 1$) containing at least a track with $P_T > 9$ GeV/c, for the *tight* or *loose* electrons;
- the number of forward jets ($1.2 < |\eta| < 2.8$) for the *plug* electrons.

The *fake rate* is evaluated on the inclusive jet samples acquired with the JET20, JET50, JET70 and JET100 trigger paths, that ask for a jet in the event with a transverse energy of at least 20, 50, 70 or 100 GeV respectively. The *fake rates*, measured in the inclusive jet samples, are shown in figure 5.12 as a function of the jet transverse energy. The average *fake rate* is parametrized with an exponential function of the jet transverse energy ($c + a \cdot e^{b \cdot E_T}$), as shown in figure 5.13, both for central and plug jets. A 50% systematic uncertainty is associated to the *fake rate* and it is evaluated from the variation of the measured fake rate over the different jet samples (different jet trigger paths).

The QCD and $W + jet$ background is then evaluated from the data sample: for events with only one central electron, all the possible combination with a jet satisfying the electron kinematic requests are considered. The combination with an electron-jet invariant mass in the 66-116 GeV/ c^2 range are used for the background evaluation with a weight given by the *fake rate* of the selected jet.

Electron and jet candidates are reconstructed according to different algorithms. This implies that a the energy of *fake electron* is smaller than the energy of the corresponding jet, as shown in figure 5.14. Therefore for every selected electron-jet pair, the jet energy need to be corrected by a factor:

$$\xi = \frac{E_{jet \text{ as a fake electron}}}{E_{jet}} = 0.89 \quad (5.27)$$

This scale factor is extracted from a gaussian fit to the distribution of the ratio of the *fake electron* energy and the corresponding jet energy (figure 5.14).

Top and di-boson background

The contribution to the background of ZZ , ZW , WW and $t\bar{t}$ processes is evaluated exploiting Monte Carlo simulations. The full selection chain is applied to the simulated samples, and the selected events are normalized to

the data integrated luminosity. The di-boson samples are generated by Alpgen+Pythia combination (section 1.4), while the top sample is generated by Pythia, assuming a *top* quark mass $m_t = 175 \text{ GeV}/c^2$. The NLO production cross section is used to normalize the di-boson and $t\bar{t}$ samples.

The uncertainties on the normalization of the backgrounds due to di-boson and $t\bar{t}$ are estimated to be $\pm 20\%$. These uncertainties include the NLO cross section indetermination estimated by varying the factorization scale by a factor two ([28]), the cross section uncertainty due to the indetermination on the top quark mass.

5.3 Muon selection

The identification of a muon candidate is based on the reconstruction of a track with an associate calorimetric deposit compatible with a *minimum ionizing particle*. Further information can be added by the matching of the track with the track segment (*stub*) reconstructed by the CMUP ($|\eta| < 0.7$) or the CMX ($0.7 < |\eta| < 1$) detector. A candidate muon with no matched *stub* in the muon system is defined as *stubless* muon. The reconstructed quantities used to select the muons are briefly described in the following.

- Track variables
 - P_T : projection on the transverse plane ($r - \phi$) of the track momentum measured by the COT;
 - $TrkAxSeg$ and $TrkStSeg$: these are the numbers of axial and stereo super-layers, respectively, with at least 5 hits.
 - d_0 : it is the impact parameter of the track in the transverse plane ($r - \phi$), corrected after the offline reconstruction of the beam line position; the resolution on the impact parameter for tracks calculated exploiting also hits recorded by the silicon vertex detector is $\sim 40\mu m$; if the track has no silicon hits attached, the request on this variable is looser;

Variable	CMUP	CMX
Cosmic	No tag	No
p_T	$> 20 \text{ GeV}/c$	$> 20 \text{ GeV}/c$
Track $ z_0 $	$< 60 \text{ cm}$	$< 60 \text{ cm}$
# COT hits	> 0	> 0
E_{Em}	$< 2 + \max(0, (p - 100)0.0115) \text{ GeV}$	$< 2 + \max(0, (p - 100)0.0115) \text{ GeV}$
E_{Had}	$< 6 + \max(0, (p - 100)0.028) \text{ GeV}$	$< 6 + \max(0, (p - 100)0.028) \text{ GeV}$
Iso/p_T	< 0.1	< 0.1
Tracknoslhits d_0	$< 0.2 \text{ cm}$	$< 0.2 \text{ cm}$
Tracksislhits d_0	$< 0.02 \text{ cm}$	$< 0.02 \text{ cm}$
TrkAxSeg	$\geq 2 \text{ SL with } 5 \text{ hits/SL}$	$\geq 2 \text{ SL with } 5 \text{ hits/SL}$
TrkStSeg	$\geq 2 \text{ SL with } 5 \text{ hits/SL}$	$\geq 2 \text{ SL with } 5 \text{ hits/SL}$
Δx_{CMU}	$< 7 \text{ cm}$	-
Δx_{CMP}	$< 5 \text{ cm}$	-
Δx_{CMX}	-	$< 6 \text{ cm}$
ρ_{COT}	-	$> 140 \text{ cm}$

Table 5.9: Summary of the cuts used to select muon candidates with a *stub* in the CMUP or CMX sub-detectors.

Variable	Stubless
Cosmic	No tag
p_T	$> 20 \text{ GeV}/c$
Track $ z_0 $	$< 60 \text{ cm}$
# COT hits	> 0
E_{Em}	$< 2 + \max(0, (p - 100)0.0115) \text{ GeV}$
E_{Had}	$< 6 + \max(0, (p - 100)0.028) \text{ GeV}$
Iso/p_T	< 0.1
$E_{\text{Em}} + E_{\text{Had}}$	> 0.1
TrkAxSeg	$\geq 2 \text{ Axial SL with } 5 \text{ hits/SL}$
TrkStSeg	$\geq 2 \text{ Sterei SL with } 5 \text{ hits/SL}$
Tracknoslhits d_0	$< 0.2 \text{ cm}$
Tracksislhits d_0	$< 0.02 \text{ cm}$

Table 5.10: Summary of the cuts used to select muon candidates reconstructed without a *stub* in the muon detectors (*stubless* muons).

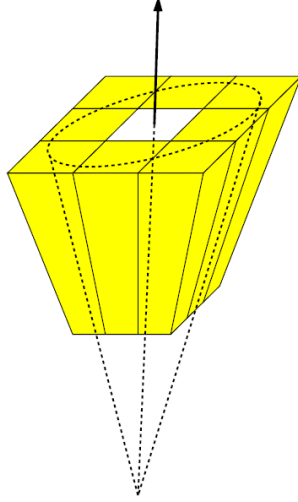


Figure 5.15: The calorimetric isolation of a candidate muon is evaluated considering the energy deposits of the towers included in a $R = 0.4$ cone.

- z_0 : it is the position along the z axis of the point of closest approach of the track to the beam line;
- $\Delta x_{CMU, CMP, CMX}$: separation between the *stub* and the track extrapolated respectively to the CMU, CMP and CMX detector plane.
- ρ_{COT} : it is the distance from the beam line at which the track crosses one of the endcap planes of the COT.

- Calorimetric variables

- E_{em}, E_{had} : energy deposited by the candidate muon in the electromagnetic (E_{em}) and hadronic (E_{had}) calorimeters. The cut applied on these variables are parametrized as a function of the track momentum.
- *Isolation* : it is a measurement of the calorimetric activity around the candidate muon. For the muon identification, the isolation is defined as:

$$Isol = \frac{1}{P_T} \left(\sum_{R < 0.4} E_T^i - E_T^\mu \right) \quad (5.28)$$

where E_T^i is the transverse energy of the i -th tower, E_T^μ is the transverse energy deposited in the tower crossed by the track and P_T is the track momentum. The sum is performed over all the towers inside a cone with a radius $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4$ around the track direction, as shown in figure 5.15. The request on isolation increases the separation between isolated muons coming from vector bosons decay and muons produced in a semileptonic decay of a hadron.

The candidate muons are divided in categories (CMUP, CMX, Stubless) according to the set of cuts they satisfy, as reported in table 5.9.

The distributions of the quantities used for the muon identification are shown in the figure from 5.16 to 5.23, compared with the corresponding MC prediction.

5.3.1 Muon trigger efficiency

The procedure used by CDF collaboration to evaluate the trigger efficiency of the high- P_T muon triggers is based on a very pure $Z \rightarrow \mu^+ \mu^-$ sample. These events are identified through the reconstruction of a pair of identified CMUP or CMX muons with invariant mass in a narrow window around the Z mass (76-106 GeV/ c^2), and with $|z_0^{(1)} - z_0^{(2)}| < 4cm$. Furthermore at least one muon must satisfy the trigger requests. The other muon is then exploited for the evaluation of the trigger efficiencies for the high- P_T trigger paths.

In the latest data taking periods the standard high- P_T CMX trigger has been substituted by a set of trigger paths with features more suitable to the increasing instantaneous luminosity provided by Tevatron. The new CMX trigger paths are dynamically prescaled or enabled only when the instantaneous luminosity decreases under certain thresholds. These features increase the uncertainty on the evaluation of the integrated luminosity acquired with the new version of the high- P_T CMX muon triggers. Nevertheless the application of these new trigger paths is limited to the last period of data taking, and the

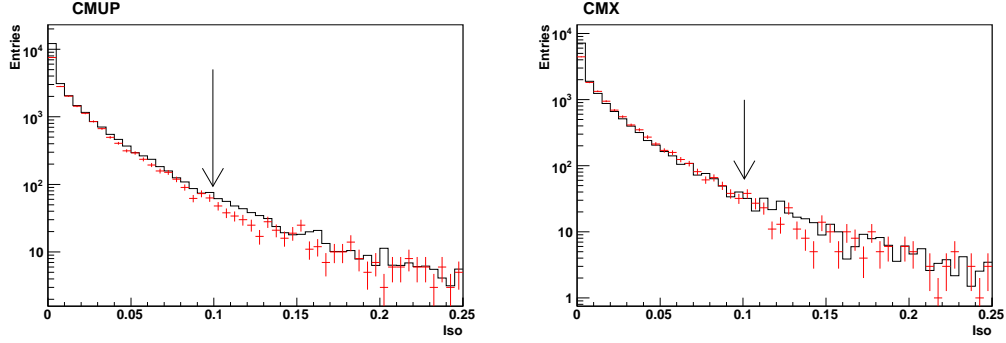


Figure 5.16: Comparison between data (red) and MC (black) of the *Isolation* distributions for CMUP (left panel) and CMX (right panel) muons. The arrows indicate the standard identification cut.

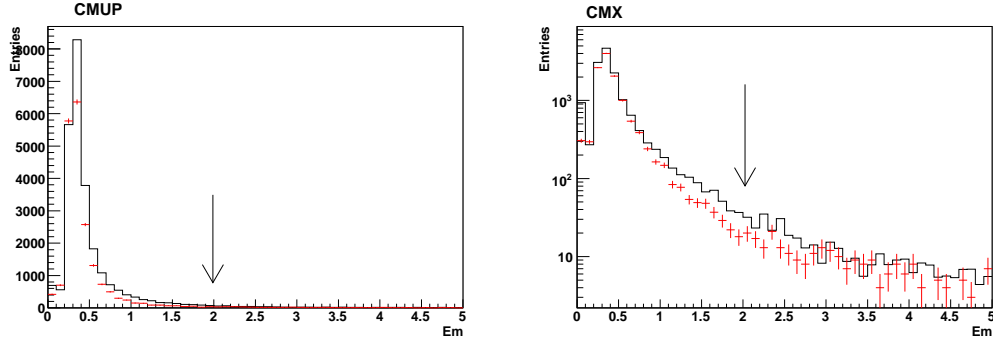


Figure 5.17: Comparison between data (red) and MC (black) of the distributions of the energy released in the electromagnetic calorimeter, for CMUP (left panel) and CMX (right panel) muons. The arrows indicate the standard identification cut.

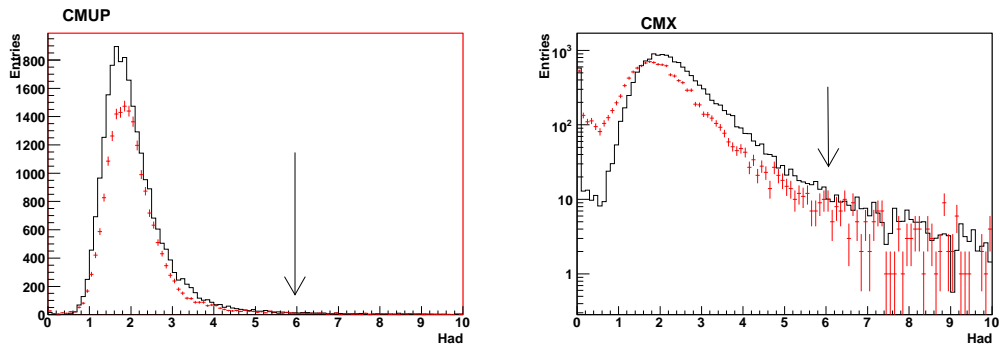


Figure 5.18: Comparison between data (red) and MC (black) of the distributions of the energy released in the hadronic calorimeter, for CMUP (left panel) and CMX (right panel) muons. The arrows indicate the standard identification cut.

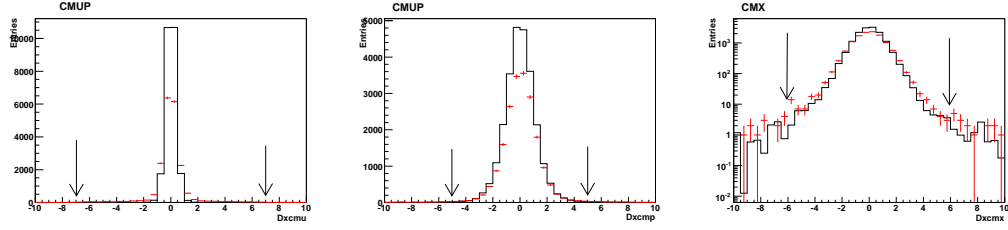


Figure 5.19: Comparison between data (red) and MC (black) of the separation between the *stub* and the track extrapolated respectively to the CMU (left panel), CMP (central panel) and CMX (right panel) detector plane. The arrows indicate the standard identification cut.

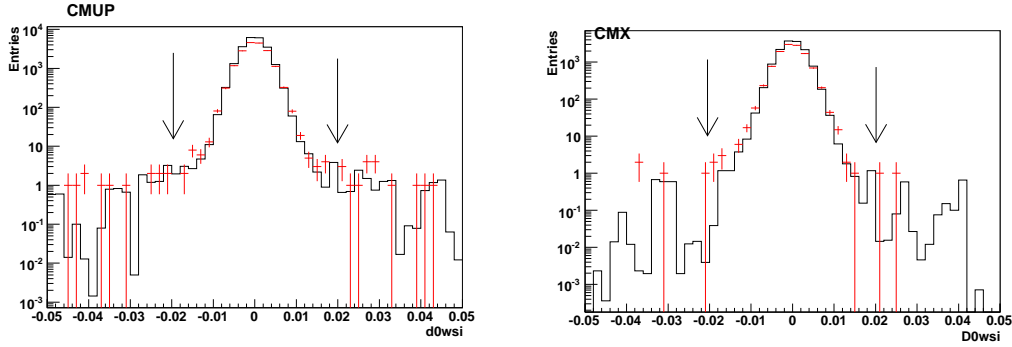


Figure 5.20: Comparison between data (red) and MC (black) of the impact parameter distributions of tracks reconstructed with COT and silicon hits, for CMUP (left panel) and CMX (right panel) muons. The arrows indicate the standard identification cut.

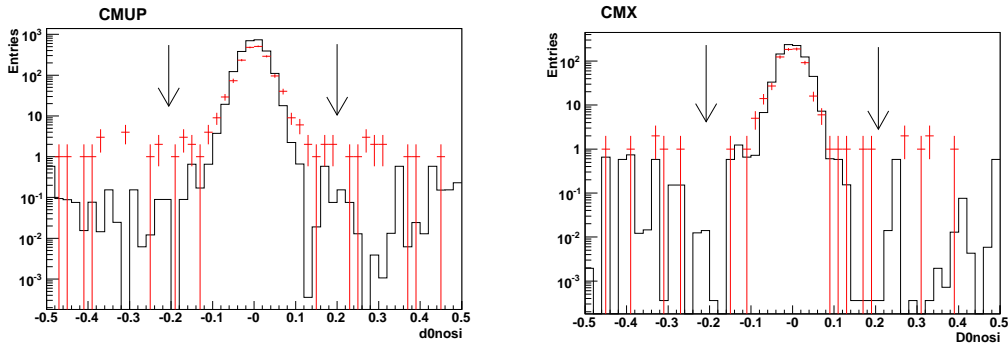


Figure 5.21: Comparison between data (red) and MC (black) of the impact parameter distributions of tracks reconstructed with no silicon hits, for CMUP (left panel) and CMX (right panel) muons. The arrows indicate the standard identification cut.

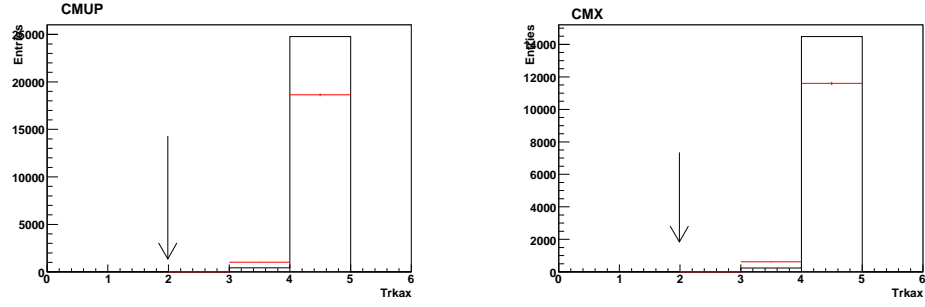


Figure 5.22: Comparison between data (red) and MC (black) of the number of axial COT super-layers with at least 5 hits, distributions for CMUP (left panel) and CMX (right panel) muons. The arrows indicate the standard identification cut.

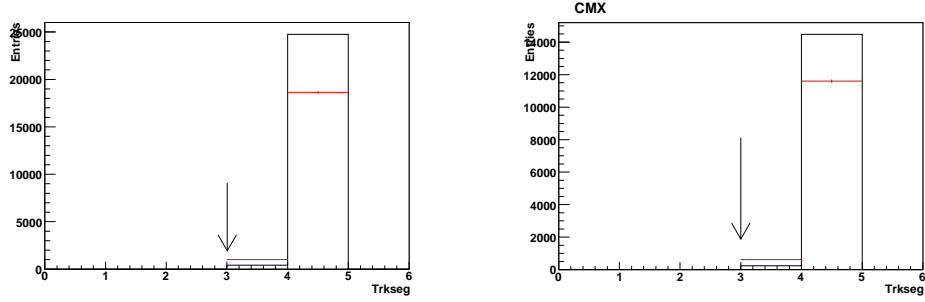


Figure 5.23: Comparison between data (red) and MC (black) of the number of stereo COT super-layers with at least 5 hits, distributions for CMUP (left panel) and CMX (right panel) muons. The arrows indicate the standard identification cut.

Trigger path	138425	190697	222526	228664
	→ 186598	→ 222426	→ 228596	→ 241664
MUON_CMUP18	0.9015	0.9173	0.9298	0.9247
MUON_CMX18	0.9665	0.9527	-	-
MUON_CMX18_L2_PT15	-	-	0.8786	-
MUON_CMX18_L2_PT15_LUMI_200	-	-	0.8097	-
MUON_CMX18_L2_LOOSE_LUMI_200	-	-	0.8520	0.8843
MUON_CMX18_LUMI_250	-	-	0.8526	0.8928

Table 5.11: Summary of the trigger efficiencies for the high- P_T muon trigger paths.

impact on the background normalization has been checked to be negligible for the present analysis.

The values of the trigger efficiency of the high- P_T muon trigger paths used in the present analysis are reported in table 5.11 according to the application run range.

5.3.2 Muon selection efficiency

The procedure used by CDF collaboration to evaluate the selection efficiency of the different muon categories is based on pure $Z \rightarrow \mu^+\mu^-$ sample. These events are identified through the reconstruction of a pair of identified muons with invariant mass in a narrow window around the Z mass (76-106 GeV/ c^2), and with $|z_0^{(1)} - z_0^{(2)}| < 4cm$. Furthermore the first muon must satisfy the CMUP or CMX requests reported in table 5.9. The other muon is then exploited for the evaluation of the selection efficiencies of the different set of selection cuts.

Any eventual disagreement between data and Monte Carlo simulation can reflect in a different value of the selection efficiencies. The standard correction procedure adopted by the CDF collaboration relies on the evaluation of *scale factors* to reconcile the selection efficiencies measured in MC simulation with the ones measured in data samples:

$$SF = \frac{\epsilon_{sel}^{Data}}{\epsilon_{sel}^{MC}} \quad (5.29)$$

The *scale factors* applied in the present analysis to correct the MC selection are reported in table 5.12.

5.3.3 $Z \rightarrow \mu^+\mu^-$ events selection

Events with a Z boson decaying in an muon pair are identified according to the following requests:

- a first $\mu^{+/-}$ satisfying CMUP or CMX cuts;
- a second $\mu^{+/-}$ satisfying CMUP, CMX or Stubless cuts;

Run range	CMUP	CMX (Arches)	CMX (Miniskirt and Keystone)	Stubless
138425 - 186598	0.9360	1.0098	-	1.0358
190697 - 203799	0.9257	0.9927	0.9159	1.0358
203819 - 222426	0.9257	0.9927	0.9159	1.0358
222529 - 228596	0.9402	0.9370	1.0065	1.0358
228664 - 241664	0.9964	0.9370	1.0038	1.0358

Table 5.12: Identification efficiency scale factors for muon categories in different run ranges. The CMX muons are divided in two subcategories (*Arches* and *Miniskirt and Keystone*), according to the sub-detector in which they are reconstructed.

- the two selected muons must have opposite charge;
- the invariant mass of the lepton pair in the 66-116 GeV/ c^2 range.

The same selection is applied to both data and MC samples. Given that the MC simulation does not include the trigger simulation, each selected MC event is weighted according to the combined trigger efficiency of the two leptons. In events in which both muons have a reconstructed stub, the combined trigger efficiency ϵ^{trig} is evaluated as

$$\epsilon^{trig} = 1 - (1 - \epsilon_1^{trig}) \cdot (1 - \epsilon_2^{trig}) \quad (5.30)$$

where $\epsilon_{1,2}^{trig}$ are the muons trigger efficiencies. In events in which only the first muon has a reconstructed stub, the trigger efficiency is given by

$$\epsilon^{trig} = \epsilon_1^{trig} \quad (5.31)$$

Also the selection efficiency of the $Z \rightarrow \mu^+\mu^-$ events can be evaluated combining the selection efficiencies of the selected leptons. Given that the muon categories are non-overlapping, the combined selection efficiency for $Z \rightarrow \mu^+\mu^-$ events is given by

$$\epsilon^{sel} = \epsilon_1^{sel} \cdot \epsilon_2^{sel} \quad (5.32)$$

and selection efficiency *scale factor* is then given by:

$$SF = SF_1 \cdot SF_2 \quad (5.33)$$

5.3.4 Background to $Z \rightarrow \mu^+\mu^-$ events

The sources of background events for the $Z \rightarrow \mu^+\mu^-$ process are the QCD or $W + jets$ events, in which one or more jets are reconstructed as electrons, and other processes with a real Z boson or in which the leptonic decay can mimic the decay of a Z boson, like the di-boson (ZZ , ZW and WW) and the $t\bar{t}$ events.

QCD and $W + jet$

Muon pairs produced in QCD and $W + jet$ events have approximately the same probability to have same or opposite charge. The background contribution from these processes, can then be evaluated asking same charge muons in the Z selection. This background component is much reduced respect to the $Z \rightarrow e^+e^-$ case. Furthermore in the muon case, the background sample is built of few events with unit weight, while in the electron case it is built of many events with very small weight. For these reasons the QCD and $W + jet$ background distributions for $Z \rightarrow \mu^+\mu^-$ events are less smooth respect to the corresponding distributions for the electron decay (see figure 5.26).

Top and di-boson background

The contribution to the background of ZZ , ZW , WW and $t\bar{t}$ processes is evaluated exploiting the MC simulation. The full selection chain is applied to the simulated samples, and the selected events are normalized to the data integrated luminosity. The di-boson samples are generated by Alpgen+Pythia combination (section 1.4), while the top sample is generated by Pythia, assuming a *top* quark mass $m_t = 175 \text{ GeV}/c^2$. The NLO production cross section is used to normalize the di-boson and $t\bar{t}$ samples.

The uncertainties on the normalization of the backgrounds due to di-boson and $t\bar{t}$ are estimated to be $\pm 20\%$. These uncertainties include the NLO cross section indetermination estimated by varying the factorization scale by a factor two ([28]), the cross section uncertainty due to the indetermination on the top quark mass.

5.4 Z boson selection results

A summary of the events selected in signal MC and in the background samples is reported in table 5.13, compared with the number of events selected in data. The uncertainty associated to the signal+background prediction is mainly due to the uncertainty on the integrated luminosity measurement (6%), and includes the contributions of the systematic uncertainty on the trigger efficiency and the uncertainties on the background estimations.

The total invariant mass distribution of the reconstructed Z bosons in both the e^+e^- and $\mu^+\mu^-$ decay channels is shown in figure 5.24. The invariant mass distributions of the reconstructed $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ events are shown in figure 5.25 and 5.26 respectively.

5.5 *Anti-top* cuts

The top events represent the main background for $Z + b$ events. In fact the top quark always decays in a b quark and a W boson ($t \rightarrow bW$). In case of leptonic decay of the W boson ($W \rightarrow l\nu$), the $t\bar{t}$ event can easily fake the $Z + b$ signature. The expected production cross section for $t\bar{t}$ events with two charged leptons in the final state is given by

$$\sigma(t\bar{t}) \cdot BR(W \rightarrow l\nu) \cdot BR(W \rightarrow l\nu) \simeq 0.08pb \quad (5.34)$$

to be compared with the production cross section predicted for $Z + b$ events, with the Z boson decaying in two leptons

$$\sigma(Z + b) \cdot BR(Z \rightarrow e^+e^-) \simeq 0.54pb. \quad (5.35)$$

The background expected from $t\bar{t}$ production is then $\sim 15\%$ of the of the $Z + b$ signal. The indetermination on the normalization of this background component is then one on the main sources of systematic uncertainty for the measurement of the $Z + b$ cross section.

Specific requests are implemented in the events selection to reduce the top background. In figure 5.27 the MC distributions of $Z + jet$ and $t\bar{t}$ events in the

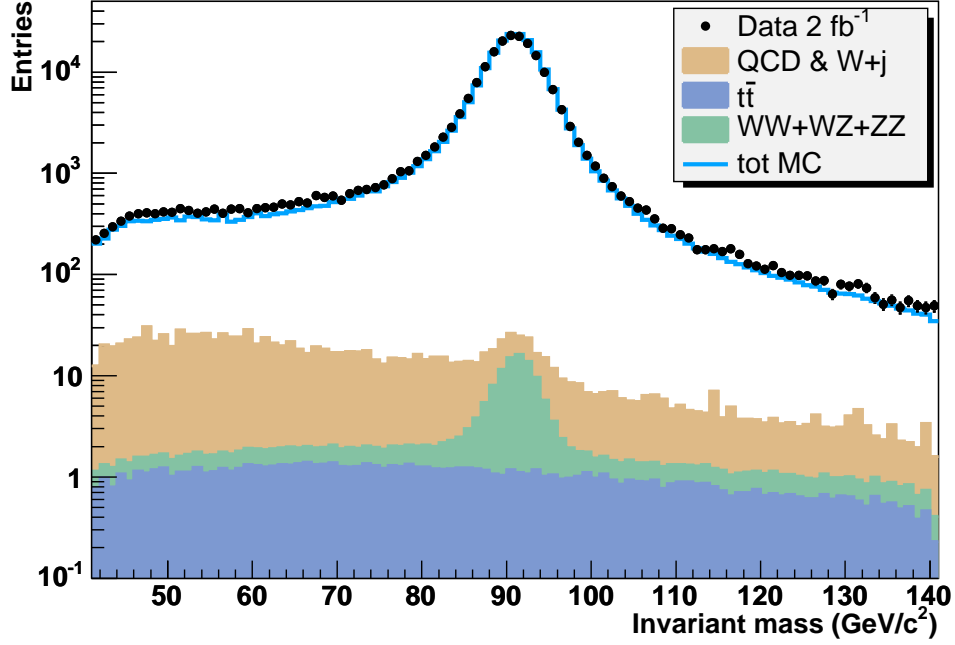


Figure 5.24: Invariant mass of the lepton pair for events in which a Z boson is reconstructed in both electron and muon decay.

Process	$Z \rightarrow e^+e^-$	$Z \rightarrow \mu^+\mu^-$	Total
QCD and $W + jet$	810 \pm 410	240 \pm 120	1050 \pm 530
$t\bar{t}$	50 \pm 10	48 \pm 10	98 \pm 20
ZZ, ZW, WW	74 \pm 15	57 \pm 11	131 \pm 26
signal MC	129500 \pm 7800	78100 \pm 4700	207600 \pm 12500
Total MC	130500 \pm 7800	78500 \pm 4700	209000 \pm 12500
Data	129196	80733	209929

Table 5.13: Number of candidate $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ events selected in signal and background MC samples. The QCD and $W + jet$ background is evaluated on data.

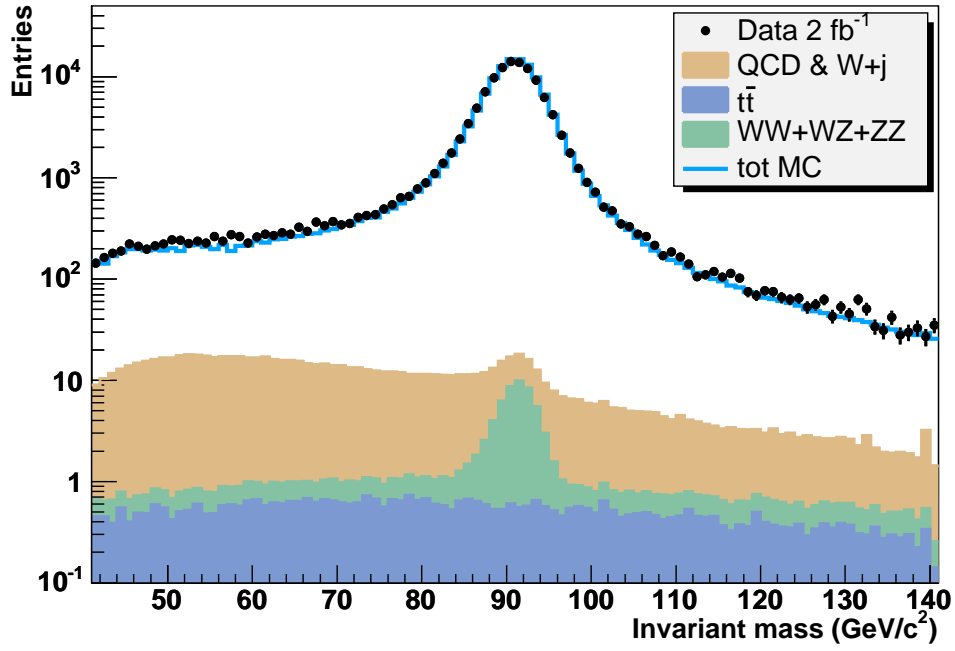


Figure 5.25: Invariant mass of the e^+e^- pair for events in which a Z boson is reconstructed by its electron decay.

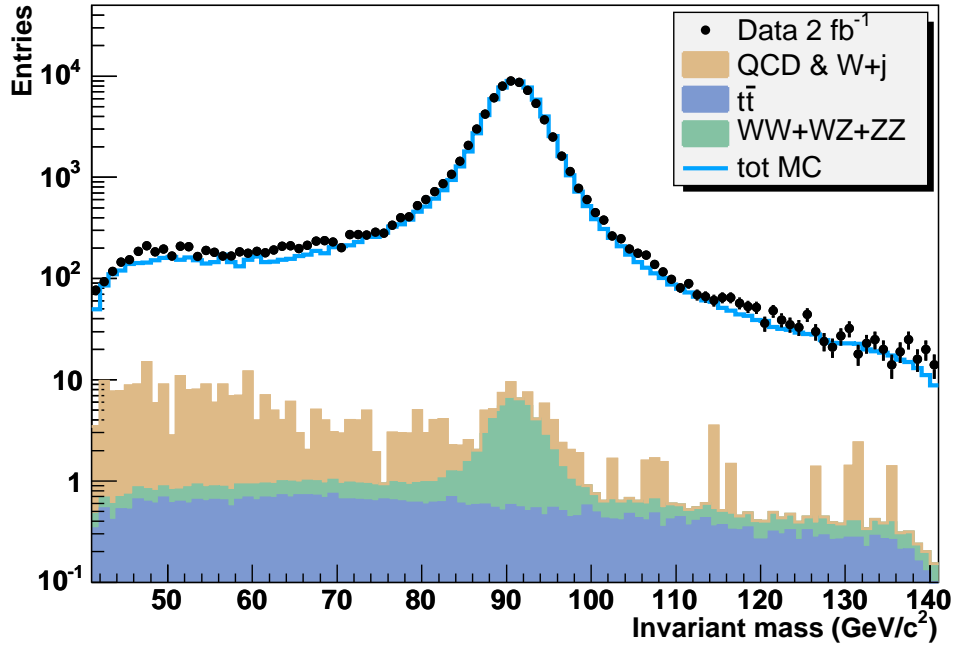


Figure 5.26: Invariant mass of the $\mu^+\mu^-$ pair for events in which a Z boson is reconstructed by its muon decay.

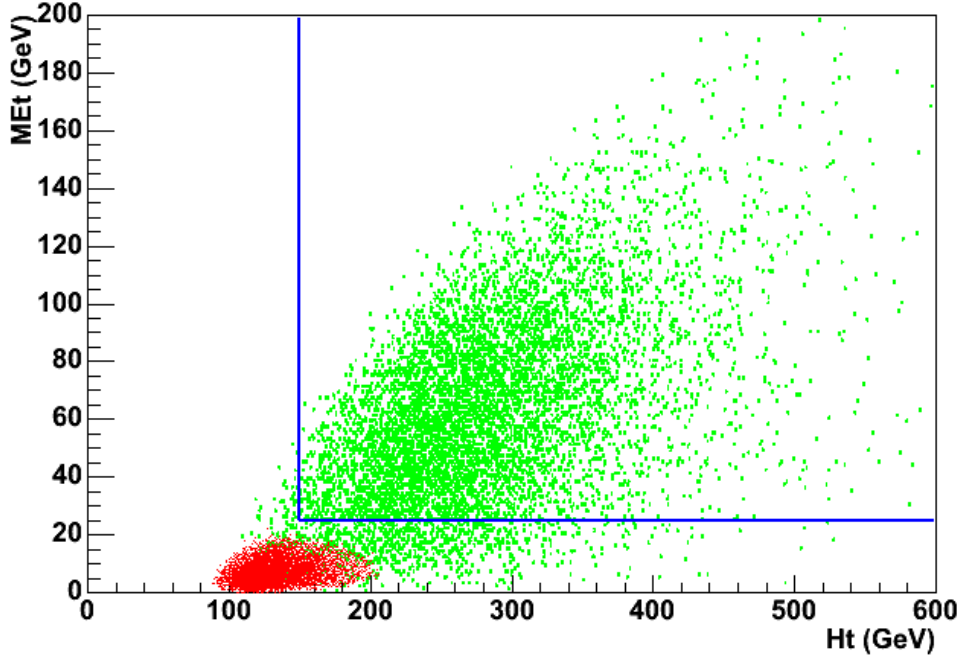


Figure 5.27: Scatter plot of the ME_T distribution versus the H_T , for $Z + jet$ (red) and $t\bar{t}$ (green) simulated events. The blue lines represent the *anti-top* cuts.

$H_T - ME_T$ plane are compared, where ME_T is the *missing transverse energy* of the event and H_T is the *scalar transverse energy*, i.e. the scalar sum of the ME_T with the transverse energy of all the electrons, muons and jets identified in the event. The $t\bar{t}$ events with two charged leptons in the final state are characterized by high values of ME_T and H_T because of the two neutrinos in the final state and the high jet multiplicity of the decay. Events are removed from the selected Z sample if $H_T > 150\text{GeV}$ and $ME_T > 25\text{GeV}$. These requirements reduce the $t\bar{t}$ background by 87.5%, whilst reducing the $Z + jet$ signal by 2%, regardless of the jet flavour.

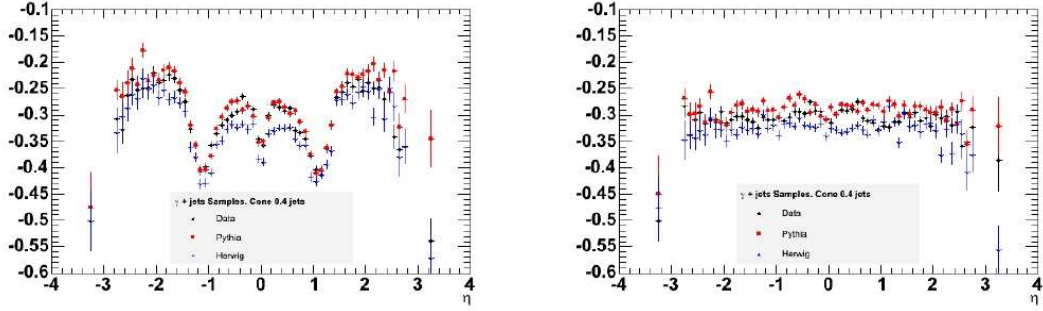


Figure 5.28: Calorimeter response to jets: uncorrected (left) and corrected (right) by di-jet balance. Events from the Pythia and Herwig generators are compared to data.

5.6 $Z + jet$ events selection

The $Z + jet$ sample is extracted from the events containing a reconstructed Z boson and satisfying the *anti-top* cuts. The hadronic jets are reconstructed using the JETCLU algorithm, described in section 1.3.1, with a $R^1 = 0.4$ cone radius.

The procedure developed by the CDF collaboration to correct the energy of the reconstructed jets for detector effects is described in section 5.6.1. In section 5.6.2 the jet selection strategy and the $Z + jet$ selected sample are presented.

5.6.1 Jet corrections

The calorimeter tower response to energy deposits is calibrated on test-beam or during the experiment using particles of known momentum.

However detector effects (for example, non linearities as e/h ratios, leakage, etc) or physics effects (for example out of cone energy, unclustered energy, etc) tend to degrade the parton four-momentum when passing to the calorimeter level. Consequently several corrections are needed.

The different corrections to jet energies are organized in levels and applied in sequence each one to the output of the previous one.

$$^1R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

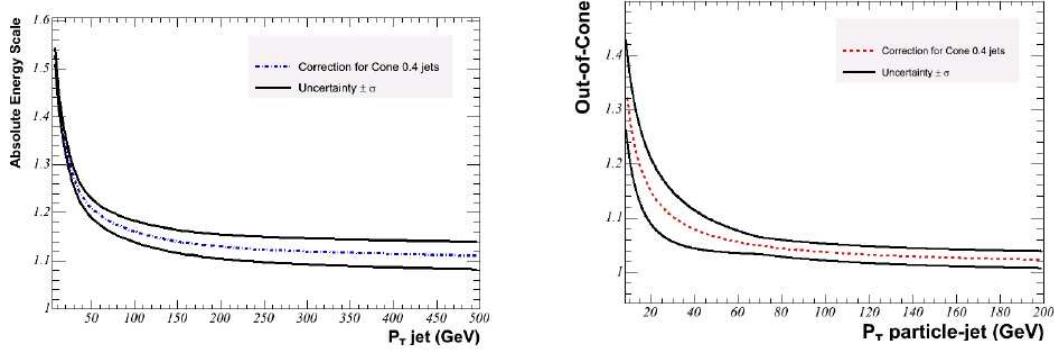


Figure 5.29: Left: absolute jet energy correction. Right: Out of Cone jet energy correction. Both corrections are plotted as functions of jet transverse momentum. The bounds are the size of the systematic error (1σ).

Level 0: these corrections are applied in the CEM to set the overall energy scale with electrons resulting from the Z^0 boson decay. The the same calibration is performed in CHA and WHA via J/ψ electrons about every 40 pb^{-1} of collected data. ^{60}Co radioactive sources and laser beams allow to transport the relative calibration to the entire calorimeter volume.

Level 1 are **relative corrections**. The η dependence of jet energy is corrected for. The differences are due to uninstrumented regions, different amount of material in the tracking volume and in the calorimeters, different responses by detector built with different technologies. The di-jet balance technique is applied. Events with exactly two jets are selected, of which one is called *trigger* and is in the region $0.2 < |\eta| < 0.6$ where the response of the calorimeter is well understood, the other one is called *probe*. The correction consists in modifying the probe jet transverse energy in order to balance the transverse energy of the trigger. A plot of the uncorrected (left) response and of the corrected response by di-jet balance (right) is reported in figure 5.28.

Levels 2 and 3 are not in use any more. Level 2 was used in Run 1 to correct for time depending variations in gain of the plug gas calorimeters. The level 3 was taking into account the differences between Run 1 and Run 2. These

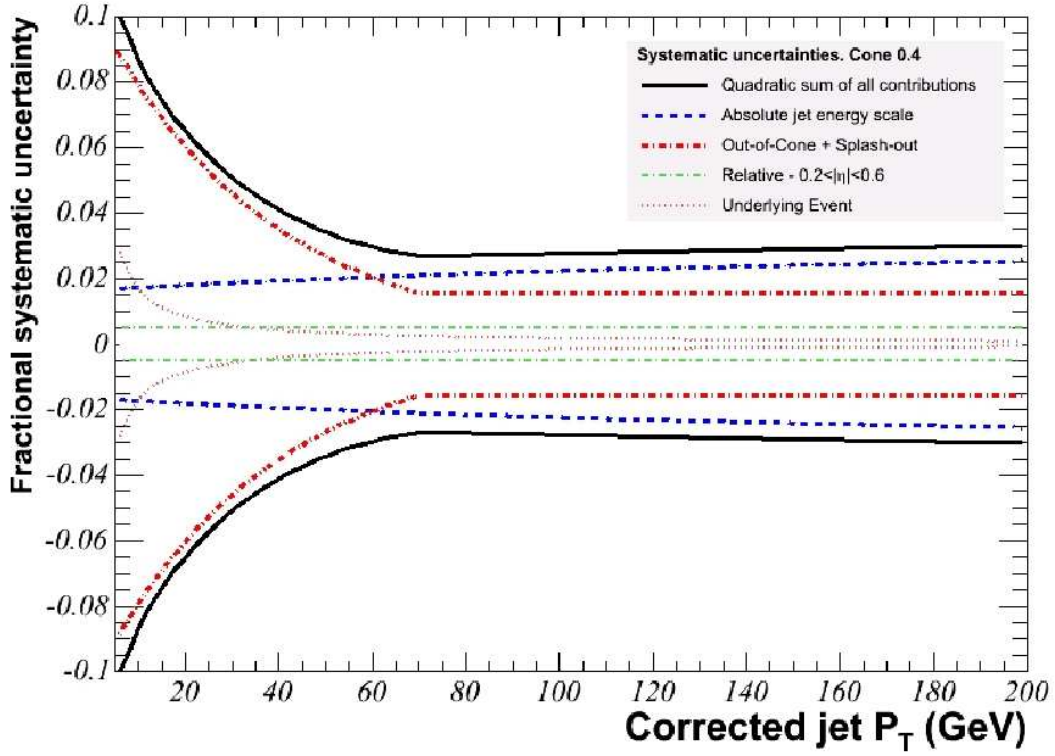


Figure 5.30: Systematic error in the jet energy measurement as the quadratic sum of the different contributions as described in the text.

differences are presently accounted for by the new generations of the jet reconstruction software.

Level 4 are **multiple $p\bar{p}$ interactions** corrections and are used to take into account the multiple $p\bar{p}$ interactions in the same bunch crossing: if more than one proton is interacting, the measured jet energy increases. The correction is derived from minimum bias data and it is parameterized as a function of the number of interaction vertexes in the recorded event.

Level 5: absolute corrections. They correct the jet energy measured in the calorimeter for non-linear response of the CEM, for energy loss due to low-energy particles that couldn't reach the calorimeter or are lost in cracks. The correction is derived by comparing the same MC events at calorimeter level and at particle level as a function of p_T . Figure 5.29,

Process	$Z(e^+e^-) + j$	$Z(\mu^+\mu^-) + j$	Total
QCD and $W + jet$	69 \pm 35	6 \pm 3.0	75 \pm 38
$t\bar{t}$	3.56 \pm 0.68	3.13 \pm 0.60	6.69 \pm 1.3
ZZ, ZW, WW	26.1 \pm 5.0	20.1 \pm 3.9	46.2 \pm 8.9
signal MC	10860 \pm 630	7570 \pm 440	18430 \pm 1100
Total MC	10960 \pm 630	7600 \pm 440	18560 \pm 1100
Data	11119	7926	19045

Table 5.14: Number of candidate $Z+jet$ events with the Z boson reconstructed in both e^+e^- and $\mu^+\mu^-$ pairs, selected in data and MC samples. The QCD and $W + jet$ background is evaluated on data.

left, shows this correction as a function of the jet transverse momentum.

Level 6: underlying event corrections are intended to remove the energy contributed by the underlying event, which is due to spectator partons and can contribute to a jet cluster.

Level 7 is the correction for the **out of cone** (OOC) energy. It adds to the jets the energy lost for leaked prongs outside the clustering cone (up to $R = 1.3$). Figure 5.29, right, shows the out of cone correction and its systematic error as a function of P_T .

Level 8: splash out corrects for the additional energy lost outside the $R = 1.3$ cone.

Figure 5.30 shows the entire set of systematic errors in measuring the jet energy. The (quadratic) sum of all contribution is shown as a black thick curve.

5.6.2 Jet selection

In events with a reconstructed Z boson the hadronic jet candidates are selected in order to provide a suitable sample for the determination of the Heavy Flavour fractions. For the present analysis the *absolute correction* (level 5 of

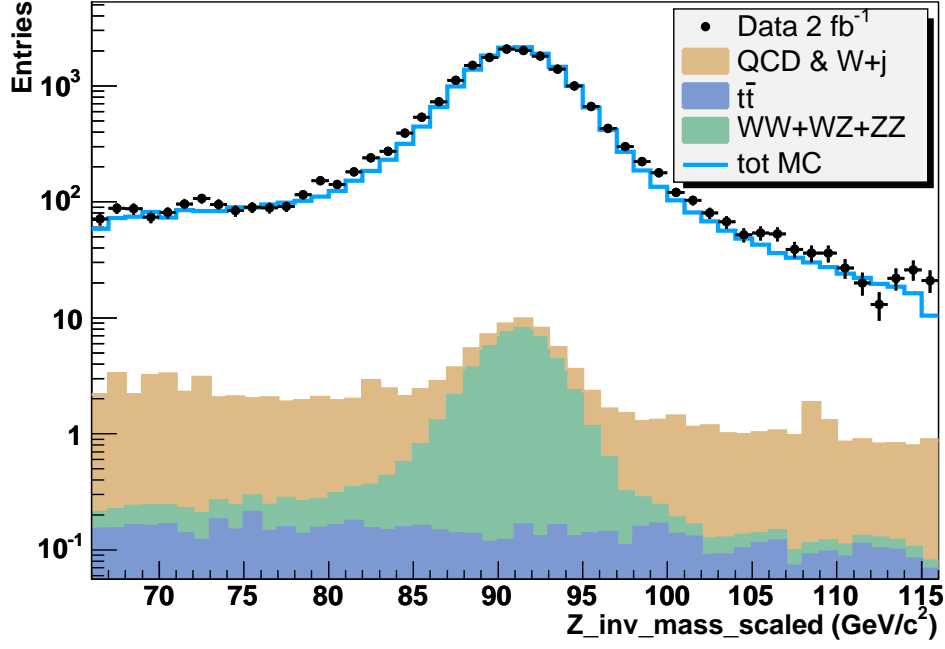


Figure 5.31: Invariant mass of the lepton pair for events with a reconstructed Z boson and at least one jet.

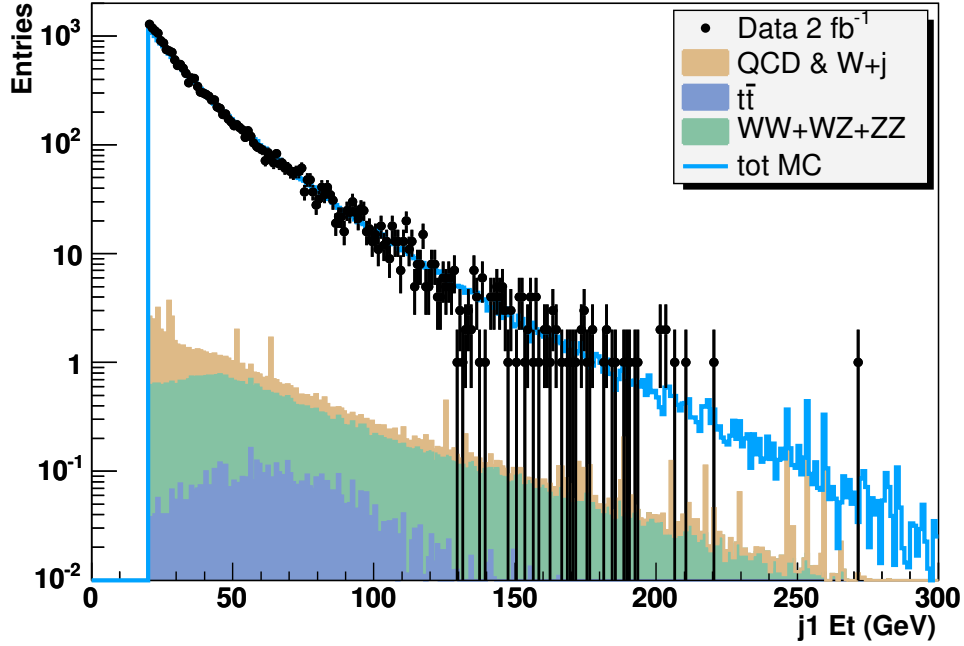


Figure 5.32: Transverse energy distribution of the *leading jet* for events with a reconstructed Z boson and at least one jet.

the jet corrections described in section 5.6.1) is applied to the jet energy. Furthermore only central jets ($|\eta| < 1.5$) are selected to exploit the high efficiency region of the tracking system for the Heavy Flavour identification.

Selected jets must satisfy the following requests:

- $E_T^{L5} > 20$ GeV;
- $|\eta| < 1.5$;
- $\Delta R(jet, lepton) > 0.4$ for each lepton identified as a Z boson decay product.

A summary of the events selected in signal MC and in the background samples is reported in table 5.14, compared with the number of events selected in data. The uncertainty associated to the signal+background prediction is mainly due to the uncertainty on the integrated luminosity measurement (6%), and includes the contributions of the systematic uncertainty on the trigger efficiency and the uncertainties on the background estimations.

The selected jet with the highest E_T^{L5} in the event is called the *leading jet*. The invariant mass distributions of the selected lepton pair for events with at least one selected jet is shown in figure 5.31. The distributions of the E_T^{L5} , η , and ϕ of the leading jet for events with a reconstructed Z boson are shown in figure 5.32, 5.33 and 5.34 respectively.

In the following sections, if not otherwise specified, the *absolute correction* is assumed for the jet transverse energy.

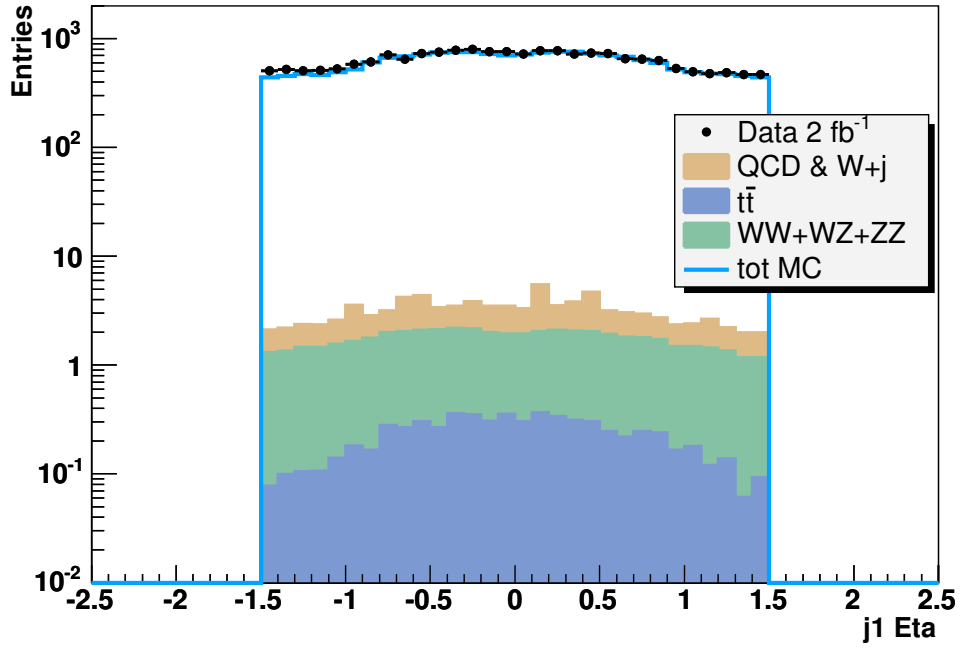


Figure 5.33: Pseudo-rapidity distribution of the *leading jet* for events with a reconstructed Z boson and at least one jet.

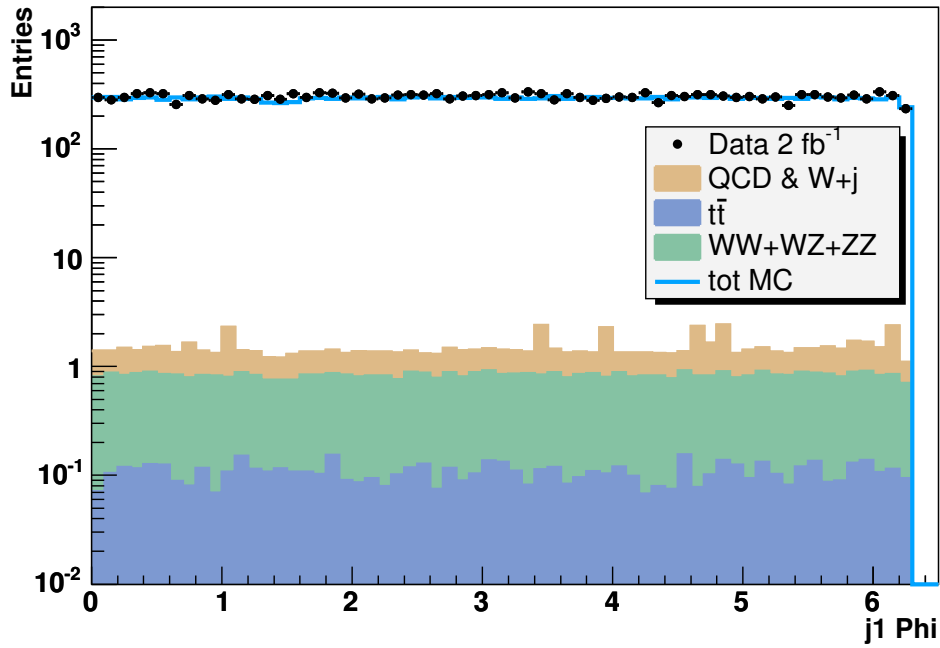


Figure 5.34: Polar angle distribution of the *leading jet* for events with a reconstructed Z boson and at least one jet.

Chapter 6

Heavy Flavour fraction measurement

The selection described in chapter 5 provides a sample of events with a Z boson produced in association with at least one jet. Exploiting the features of the new Neural Network tagger it is possible to determine the flavour composition of the leading jet distribution through a binned Likelihood fit to the tagger spectra of the data. The information extracted by the fit procedure is then used to calculate the production cross section of events with a b leading jet produced in association with a Z boson in $p\bar{p}$ collisions.

The procedure for deriving MC templates suitable to perform the measurement is presented in section 6.1. The fitting procedure to extract the flavour composition of the selected jet sample is presented in section 6.2. Due to the dependence of the detector and tagger efficiencies on the jet flavour, the flavour fractions measured by the fit procedure are different from the ones produced in the $p\bar{p}$ interaction. The corrections applied to calculate the Heavy Flavour fractions at production level and the result of the measurement are presented in section 6.3. A review of the systematic errors is presented in sections 6.4. Finally future improvements of this study and application to a wider range of analysis of the developed techniques are briefly discussed in section 6.5.

6.1 Neural Network distribution for $Z+jets$ events

In figure 6.1 is reported the Neural Network output distribution for leading jets in events with a reconstructed Z boson. The data distribution is compared with the Alpgen MC prediction, showing a good agreement between data and simulation.

The Neural Network output distribution for background events is evaluated on MC for di-boson and top samples, while it is extracted directly from data for QCD and $W + jet$ events, as described in chapter 5. This background component is much reduced in muon decay channel, with respect to the electron one. Furthermore, as described in section 5.3.4, the fake muon distributions are estimated through a different technique with respect to the electron ones. The background composition is reported in table 6.1. The application of the anti-top cuts, described in section 5.5, ensures that the selected data sample is largely dominated by the $Z + jets$ signal even in the region of high values of the Neural Network output, where the $Z + b$ events gather.

6.1.1 Identification Efficiency

The identification efficiency is defined as the fraction of the selected jets with a defined output for the Neural Network tagger. Possible differences in identification efficiency between data and MC affect the normalization of the MC-based background estimations (top and di-boson), which need to be

$Z + j$	di-boson	top	QCD and $W + jet$	data	Alpgen prediction
$Z \rightarrow e^+e^-$	19.5	3.0	26.1	7931	7846.5
$Z \rightarrow \mu^+\mu^-$	15.4	2.7	4	5628	5466.9
total	34.9	5.7	30.1	13559	13313.4

Table 6.1: Background composition for identifiable leading jet reconstructed in $Z + jet$ events, divided by background source and Z boson decay channel. In the last columns are reported the events yields respectively in data and Alpgen MC sample.

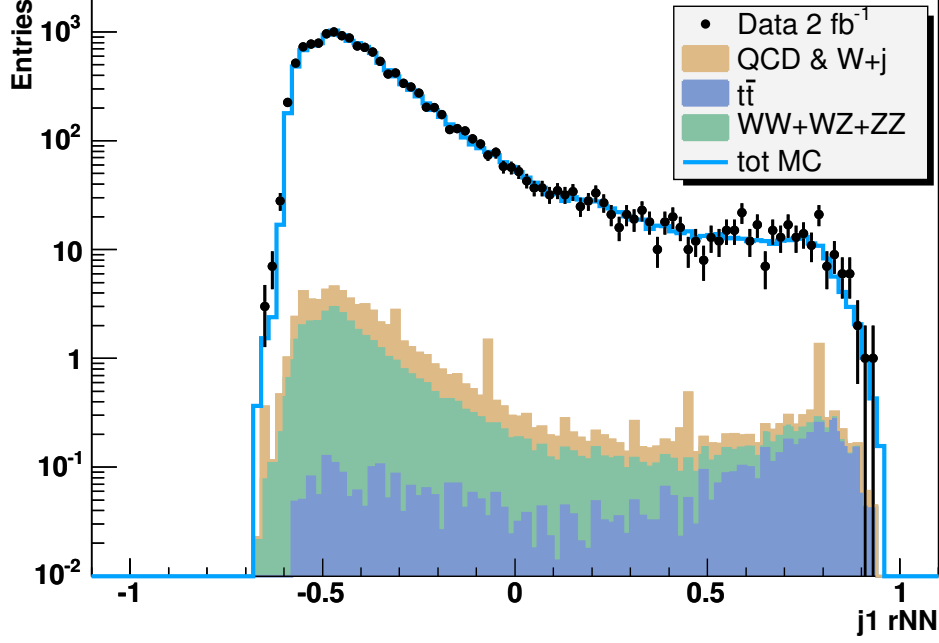


Figure 6.1: Comparison of the leading jet Neural Network output distribution for $Z + jet$ events with the MC prediction. The *spikes* in the Neural Network output distribution for QCD and $W + jet$ background events are due to the the fake muons contribution.

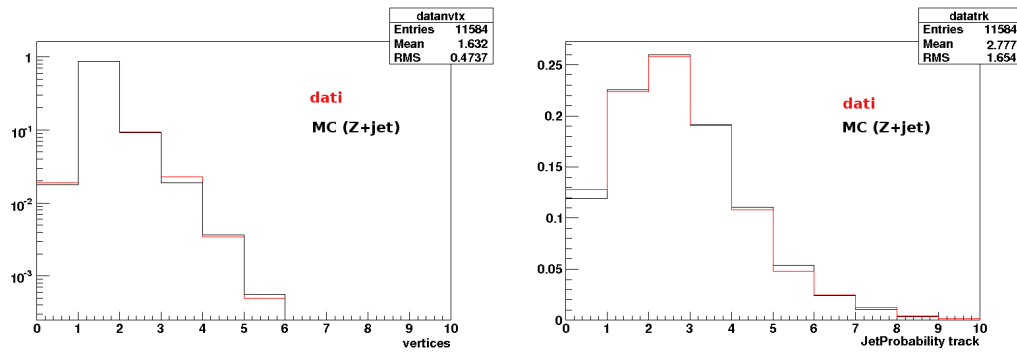


Figure 6.2: Normalized distributions of number of reconstructed vertexes (left) and number of JetProbability tracks (right) for $Z + jet$ data, compared with Alpgen MC.

corrected by a scale factor (SF_{ie}). This scale factor is evaluated both on calibration di-jet samples and on Z +jet sample, and found to be 0.96 in both cases.

As presented in section 3.2, a jet is defined *identifiable* if it has at least either one vertex reconstructed by the new vertexing algorithm or 2 tracks marked good by the JetProbability tagger. Data and MC distributions of these two variables are reported in figure 6.2 for the Z +jet sample. The comparison shows a good agreement between data and simulation except for low values of the JetProbability good tracks distribution.

Given that the definition of identifiable jet is based on the number of vertexes and tracks distributed inside the jet cone, it is expected an higher rate of identified jets among the $Z + hf$ events, respect to $Z + light - jet$ sample. The Heavy Flavour fractions in the identifiable jet sample are then altered respect to their values in the selected $Z + jet$ sample. A technique to correct the measured fractions for this effect is presented in section 6.3.1.

6.1.2 Neural Network Monte Carlo templates

The Neural Network distributions for pure flavour jet (*templates*) can be extracted from Alpgen MC samples, after applying the whole selection chain. As described in chapter 4 an appropriate procedure is adopted to avoid the double counting of events with Heavy Flavour jets. In particular events with $b - jets$ are removed from the $Z + c\bar{c}$ and $Z + p$ samples, and events with $c - jets$ are removed from the $Z + p$ sample. This procedure assures that the process of associated production of the Z boson and Heavy Flavour quark has been evaluated in the Matrix Element calculation performed by Alpgen, and not in the parton showering simulation performed by Pythia.

Nevertheless it is possible to find in the $Z + hf$ samples events in which the reconstructed leading jet is originated by a quark produced by the parton showering. To build templates suitable for the fraction fitting procedure, the simulated events have to be categorized according to the leading jet flavour. In table 6.2 the different contributions to the Neural Network templates are

source	$b - template$	$c - template$	$light - template$
$Z + b\bar{b}$	100%	1.8%	0.5%
$Z + c\bar{c}$	-	98.2%	1.6%
$Z + p$	-	-	97.9%

Table 6.2: Alpgen MC template composition, divided by source.

summarized. The templates are then built up adding all the available contributions according to the production cross sections of the different sources as evaluated by Alpgen.

6.1.3 Correction functions for $Z+jets$ events

The MC templates need to be corrected for data-MC disagreement by suitable correction functions, as described in section 4.3. The calibration procedure has been developed exploiting the high-statistic jet samples (trigger paths JET20 and JET50), and the corresponding MC samples. These calibration samples have been re-weighted in order to reproduce the same E_T spectra predicted by Alpgen MC for $Z + jet$ events. This procedure provides a good agreement of the distributions of the Neural Network tagger input variables between calibration and $Z + jet$ samples, as shown in figure 6.3 for the number of selected tracks in the event. A full calibration is performed with the re-weighted samples to extract a set of correction functions suitable for the $Z + jet$ sample.

The accuracy of the extrapolation of the calibration samples depends on the knowledge of the E_T distribution of the b , c and $light$ components of the $Z + jet$ sample. In particular the E_T distributions of $Z + b$ and $Z + c$ events have not been measured yet, and can only be extracted from MC simulation. This indetermination on the E_T distribution of $Z + hf$ events introduces a systematic uncertainty on the determination of the correction functions. The evaluation of this effect is described in section 6.4.2.

The extracted correction functions are summarized in table 6.3 and in figure

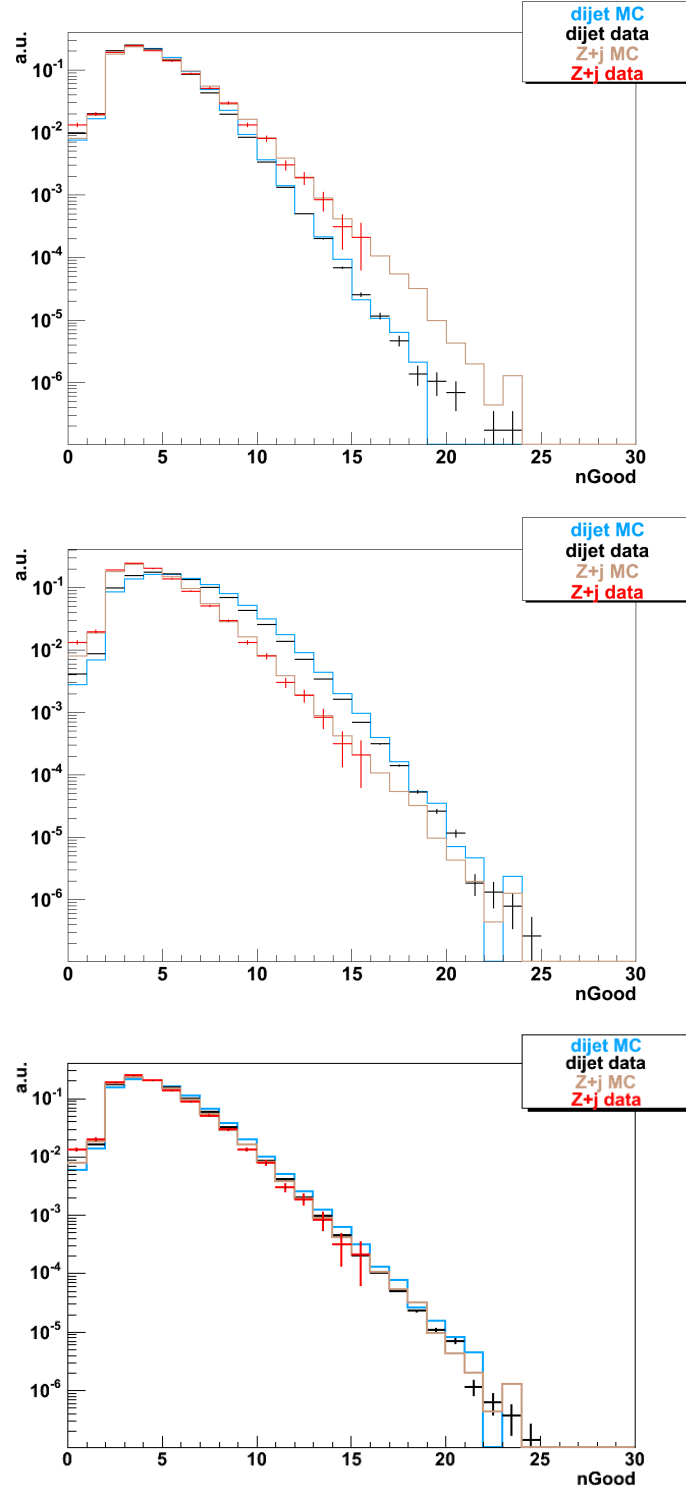


Figure 6.3: Data-MC comparison of the normalized distributions of the number of selected tracks in the event for calibration $di-jet$ sample JET20 (upper panel), JET50 (central panel) and JET20+JET50 re-weighted to the E_T distribution predicted by Alpgen simulation for $Z+jet$ events (lower panel).

q	p_0^q	p_1^q
b	1.061 ± 0.014	-0.176 ± 0.041
c	1.174 ± 0.011	1.084 ± 0.067
light	1.2678 ± 0.0014	0.6844 ± 0.0035

Table 6.3: *Correction function* parameters obtained by the calibration for the $Z + jet$ sample. Linear *correction functions* are assumed: $\phi_{jk} = p_{0,j} + p_{1,j} \cdot x_k$.

6.4 is reported how the $Z + jet$ Alpgen MC templates are modified by the calibration procedure.

6.2 Heavy Flavour fraction fit

The selection of events with a reconstructed $Z \rightarrow e^+e^-$ or $Z \rightarrow \mu^+\mu^-$ in the $[66, 116]$ GeV/ c^2 mass window, and at least one jet with $E_T > 20$ GeV and $|\eta| < 1.5$ produces a sample of 19045 events. Among them 13599 have a Neural Network output defined for the leading jet. The jet flavour composition of this sample is determined by a binned Likelihood fit analogous to the one described in section A, modified to take into account the physics and instrumental background. The inputs to the fitting procedure are Neural Network output distribution of data, the three templates (corrected as described in section 6.1.3), and the Neural Network output distribution for the background. The Negative Log Likelihood is then minimized keeping the background fraction constant. The result of the fit is reported in table 6.4.

In figure 6.5 the Neural Network output distribution for data is compared with the analogous distribution for MC, obtained adding the background spectra and the templates scaled according to measured fractions.

In order to check that the fitting procedure used is suitable to measure such small fractions a toyMC simulation is performed. Pseudo-data samples are built from the Alpgen templates, corrected for the Neural Network tagger calibration, according to the measured fractions and to the data yield. In figure 6.7 are reported the distribution of the measured fractions and the

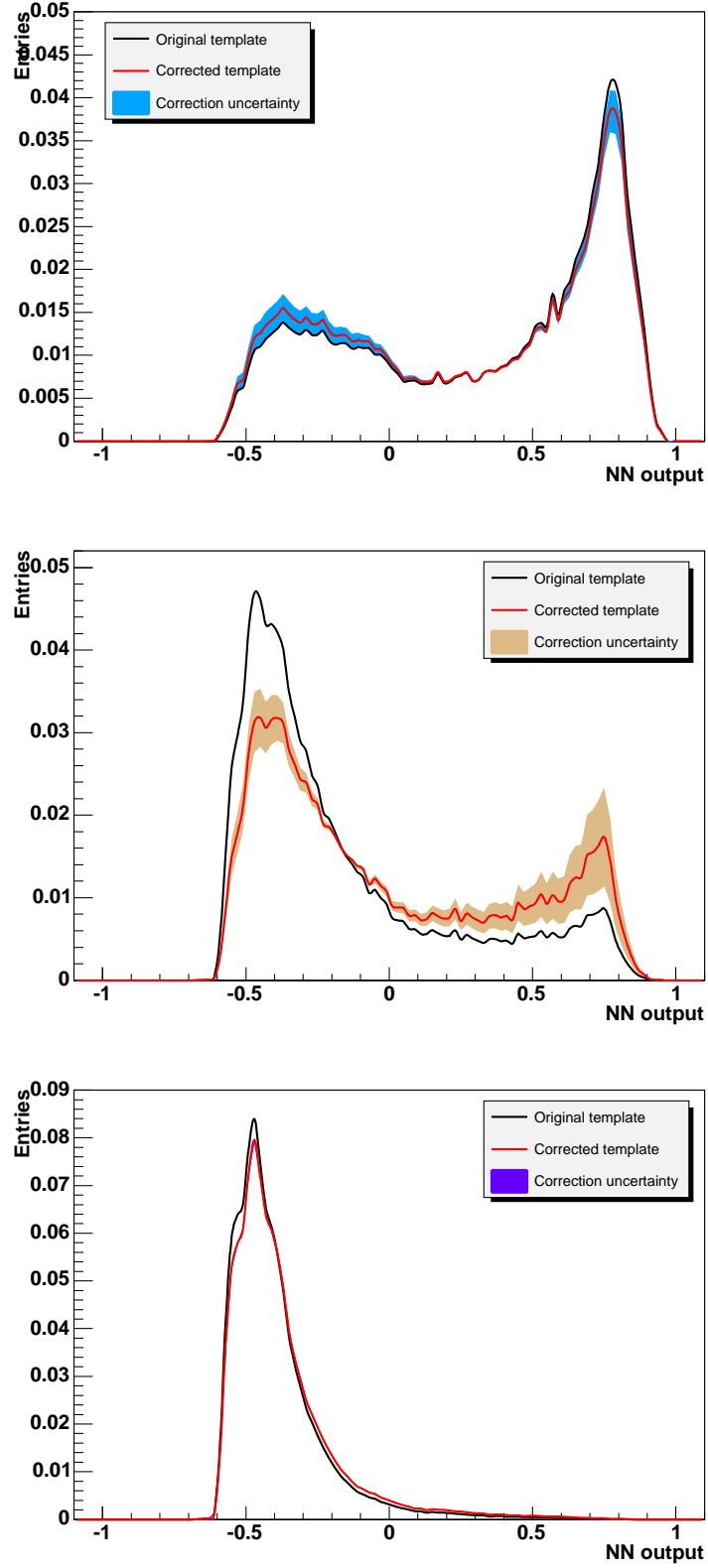


Figure 6.4: Neural Network output distribution of b (upper panel), c (central panel) and $light$ (lower panel) jets, modified by the corresponding *correction functions*, with the original MC simulation superimposed.

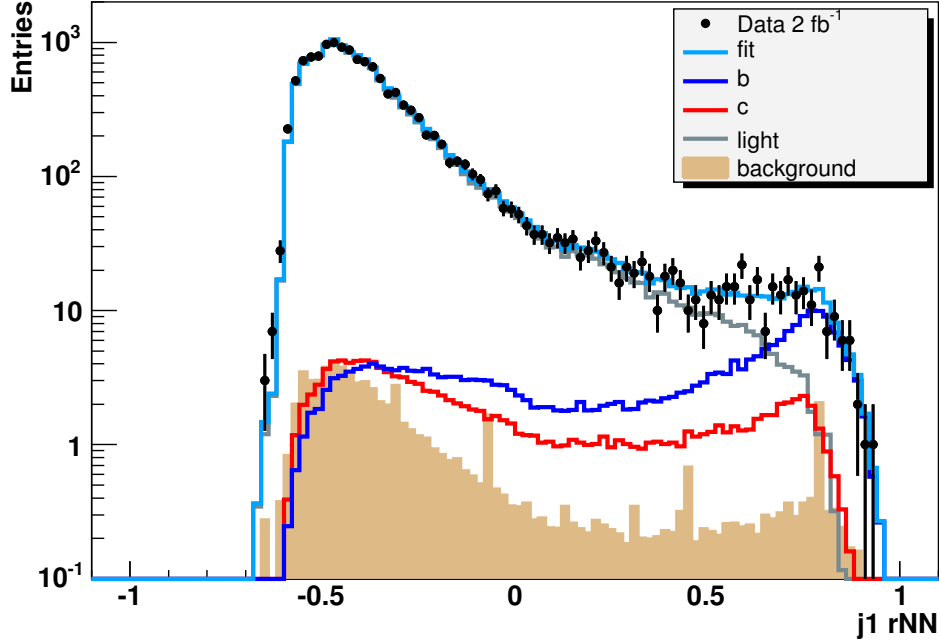


Figure 6.5: Neural Network output distribution for data compared with MC. The MC signal distribution is obtained scaling the corrected templates according to measured fractions.

corresponding errors.

The toyMC study shows that with the current statistics and the current performances of the Neural Network tagger an unbiased fit of the b fraction is achieved. With the present statistics the fit returns on average the correct value of the b fraction irrespective of the true value of the b and c fractions

Fraction	Fit	Alpgen prediction
b	0.0192 ± 0.0042	0.0115
c	0.0099 ± 0.0090	0.0188
light	0.9709 ± 0.0058	0.9697
χ^2/dof	87.7/80	89.1/80

Table 6.4: Result of the fit to data. Fitted fractions are reported with their statistical error.

correlation	value
$\rho(f_b, f_c)$	-0.852
$\rho(f_b, f_{light})$	0.587
$\rho(f_c, f_{light})$	-0.924

Table 6.5: Correlation coefficients of the fitted fractions.

in a wide range of the possible fractions combinations, as shown in figure 6.6. In addition the MINUIT estimate of the statistical uncertainty is correct since the pull distribution for the b fraction fit parameter has unit width, as shown in figure 6.7 upper left panel. On the other hand the same study shows that the c fraction parameter has a bigger statistical uncertainty, due to the lower separation power of the tagger. For this reason, and for the low c fraction determined from the fit on the data, in a 14% fraction of the generated pseudo-experiments the fit fails since the c fraction parameter hits the physical bound $f_c \geq 0$. As shown in figure 6.7 center panel, it's reasonable to assume that when the fit does not fail it behaves correctly giving an unbiased estimate of the c fraction as well with uncertainty close to be gaussian. The measurement of the c fraction is however still only marginally significant and further tests and a dedicated analysis is required to give an estimate of the allowed interval at 68% CL with correct statistical properties for the c fraction or, in alternative, to give an upper limit for the c jet production cross section. This item is left to a future revision of the present analysis while full results for the b jet production cross section and an estimate of the systematic uncertainties for both the b jet and c jet cross section ratios are quoted in the present work.

The mass range of the Z boson in the Alpgen samples used to build the templates ($Z+b\bar{b}$, $Z+c\bar{c}$, $Z+p$) is $[75, 105]$ GeV/ c^2 , while in the data samples the Z bosons are selected in the wider mass range $[66, 116]$ GeV/ c^2 . To check that the disagreement in the Z mass range does not affect the measured fractions, the MC templates are splitted in *central* (Z mass in the $[85, 95]$ GeV/ c^2 range) and *not-central* (Z mass outside the $[85, 95]$ GeV/ c^2 range), and the full fit procedure is repeated with the two categories of templates. The fractions

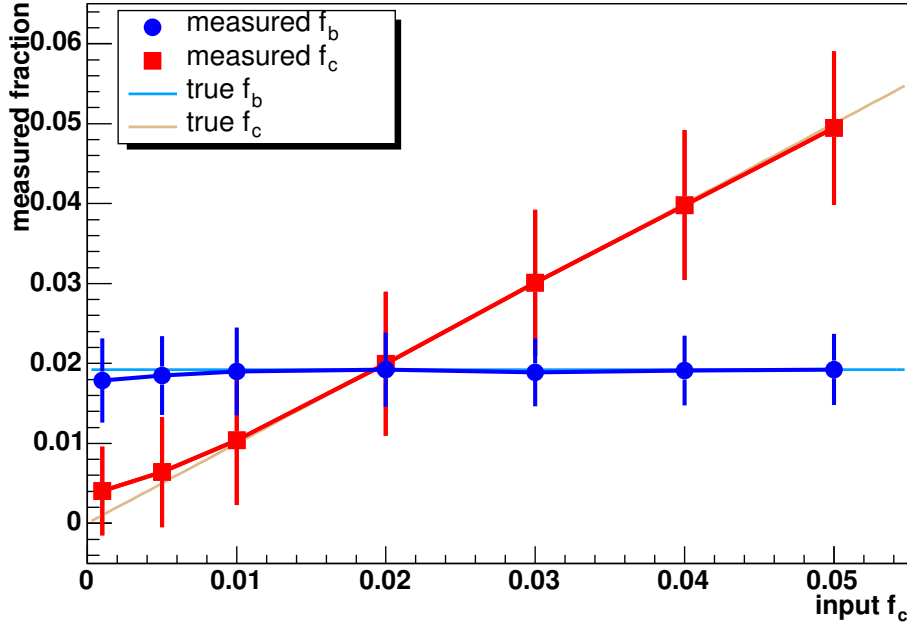


Figure 6.6: Fractions measured for b and c component by the binned Likelihood fit as a function of the true f_c . The estimate is performed through a toy MC simulation of 1000 pseudo-experiments for each configuration. The c fraction is varied from 0.001 to 0.05, while the b fraction is kept constant.

measured with the *central* and *not-central* templates agree with the result obtained with the combined templates within 10% of the statistical error for the b fraction, and within 50% of the statistical error for the c fraction.

6.3 Measurement of the cross section

The efficiencies of all the selections applied to the triggered data sample in order to extract $Z + jet$ events suitable for the fraction fit procedure, are in principle flavour dependent. This implies that the flavour composition of the fitted events differs from the flavour composition of the leading jet produced in association with a Z boson in the $p\bar{p}$ interactions. A series of corrections is applied to the measured fractions to calculate the production cross section of b leading jet in association with a Z boson.

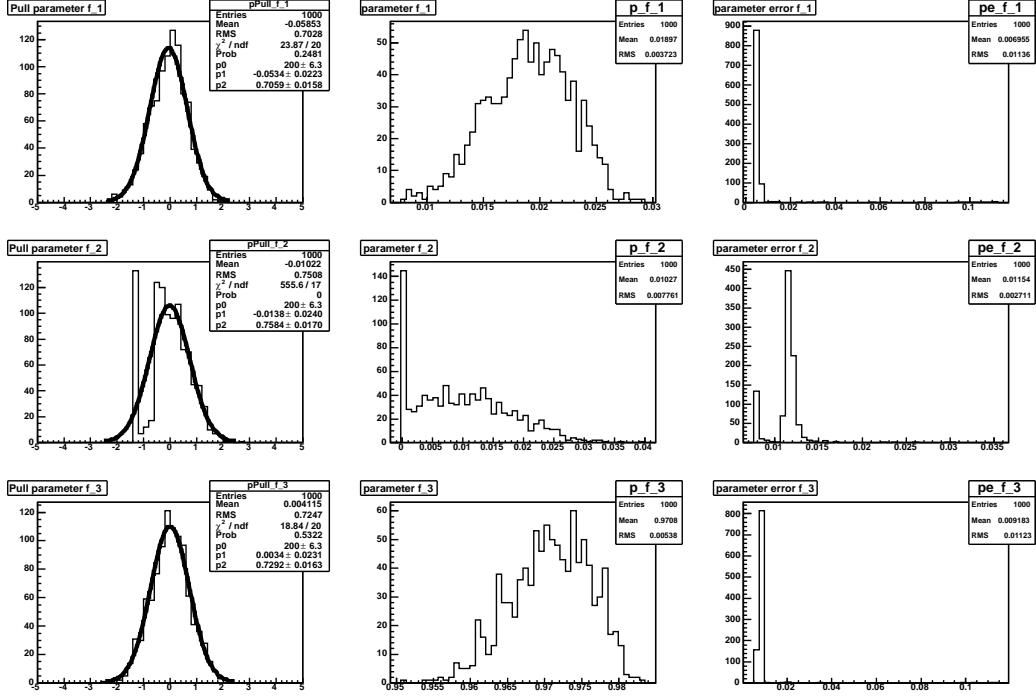


Figure 6.7: Toy MC simulation of the binned Likelihood fit procedure performed with the measured fractions as input values. For each fraction are reported the pull, fraction and fraction error distribution. The distributions are reported for b (upper panel), c (central panel) and $light$ (lower panel) fractions.

6.3.1 Identification efficiency correction

A dependence of the identification efficiency on the jet flavour modifies the fractions of Heavy Flavours in the identifiable sample, respect to the selected sample. In fact, for a given flavour q :

$$N_q^m = T_q \cdot N_q \quad (6.1)$$

where N_q^m is the number of events with a identifiable leading jet from q flavour, N_q is the number of selected events with a leading jet from q flavour, and T_q is the identification efficiency for $q-jets$, evaluated on MC samples. Expressing the number of events in terms of fractions:

$$f_q^m \cdot N^m = T_q(f_q \cdot N) \quad (6.2)$$

	T_q	f_q^m	C_q^T	f_q
$b - jets$	0.845	0.0192	0.890	0.0171
$c - jets$	0.809	0.0099	0.930	0.0092
$light - jets$	0.750	0.9709	1.003	0.9737

Table 6.6: Values of identification efficiency and correction factors, divided by jet flavour.

where f_q is the value of the fraction of selected $q - jets$, f_q^m is the fitted value for the fraction of identifiable $q - jets$, N is the total number of selected events, and N^m is the total number of selected events with a identifiable leading jet. The 6.2 can be rewritten as

$$f_q \cdot N = N^m \cdot \frac{f_q^m}{T_q} \quad (6.3)$$

Summing over the fractions:

$$N = N^m \cdot \left(\frac{f_b^m}{T_b} + \frac{f_c^m}{T_c} + \frac{f_l^m}{T_l} \right) \quad (6.4)$$

Combining the 6.3 with 6.4 it is possible to extract a correction factor (C_q^T) to correct back the fitted fractions to the selection level as:

$$f_q = C_q^T \cdot f_q^m = \frac{\frac{1}{T_q}}{\frac{f_b^m}{T_b} + \frac{f_c^m}{T_c} + \frac{f_l^m}{T_l}} \cdot f_q^m \quad (6.5)$$

In table 6.6 the values of identification efficiency and correction factors are reported as a function of the jet flavour.

6.3.2 Detector unfolding

The detector response to b and c jets may differ respect to $light$ jets in energy and rapidity resolution, energy scale and reconstruction efficiency. To measure the ratio of the cross sections for the production of events $Z + b jet$ and $Z + jet$, with the leading jet in the acceptance region $E_T > 20 GeV$, $|\eta| < 1.5$, the fractions evaluated by the fit to the Neural Network tagger spectra need

to be corrected back to the particle level, removing the detector effects. To this purpose the MC simulation is exploited:

$$\frac{\sigma(Z + b \text{ jet})}{\sigma(Z + \text{jet})} = \frac{N_{Z+b \text{ jet}}}{N_{Z+\text{jet}}} \cdot \frac{\epsilon_{sel}(Z + \text{jet})}{\epsilon_{sel}(Z + b \text{ jet})} \cdot \frac{\epsilon_j(Z + \text{jet})}{\epsilon_j(Z + b \text{ jet})} \quad (6.6)$$

with

$$\frac{N_{Z+b \text{ jet}}}{N_{Z+\text{jet}}} = f_b^m \cdot C_b^T \quad (6.7)$$

where f_b^m is the fitted b fraction and C_b^T is the identification efficiency correction factor defined in section 6.3.1. The selection efficiencies are defined as follows:

$$\epsilon_{sel}(Z + \text{jet}) = \frac{N_{rec}(Z + \text{jet})}{N_{gen}(Z + \text{jet})} \quad (6.8)$$

$$\epsilon_{sel}(Z + b \text{ jet}) = \frac{N_{rec}(Z + b \text{ jet})}{N_{gen}(Z + b \text{ jet})} \quad (6.9)$$

where $N_{gen}(Z + (b) \text{ jet})$ and $N_{rec}(Z + (b) \text{ jet})$ are respectively the number of events generated and reconstructed with at least one (b) jet in the acceptance.

The jet reconstruction efficiencies are defined as follows:

$$\epsilon_j(Z + \text{jet}) = \frac{N_{calo}(Z + \text{jet})}{N_{part}(Z + \text{jet})} \quad (6.10)$$

$$\epsilon_j(Z + b \text{ jet}) = \frac{N_{calo}(Z + b \text{ jet})}{N_{part}(Z + b \text{ jet})} \quad (6.11)$$

where $N_{calo}(Z + (b) \text{ jet})$ and $N_{part}(Z + (b) \text{ jet})$ are the number of reconstructed events with at least one (b) jet in the acceptance respectively at calorimeter and at particle level.

Furthermore the fitted fractions allow to evaluate the ratio of the cross sections for the production of events $Z + b \text{ jet}$ and Z :

$$\frac{\sigma(Z + b \text{ jet})}{\sigma(Z)} = \frac{N_{Z+b \text{ jet}}}{N_Z} \cdot \frac{\epsilon_{sel}(Z)}{\epsilon_{sel}(Z + b \text{ jet})} \cdot \frac{1}{\epsilon_j(Z + b \text{ jet})} \quad (6.12)$$

with

$$\frac{N_{Z+b \text{ jet}}}{N_Z} = \frac{N_{Z+b \text{ jet}}}{N_{Z+\text{jet}}} \cdot \frac{N_{Z+\text{jet}}}{N_Z} = f_b^m \cdot C_b^T \cdot \frac{N_{Z+\text{jet}}}{N_Z} \quad (6.13)$$

where N_Z is the number of selected events with a Z boson in the acceptance, and $N_{Z+\text{jet}}$ is the number of selected events with a Z boson and at least one jet in the acceptance.

	$Z + b\ jet$	$Z + c\ jet$	$Z + jet$
ϵ_{sel}	0.259	0.258	0.261
ϵ_j	0.849	0.943	0.998

Table 6.7: Selection and jet reconstruction efficiencies.

The selection efficiency for events with a Z boson is defined as follows:

$$\epsilon_{sel}(Z) = \frac{N_{rec}(Z)}{N_{gen}(Z)} \quad (6.14)$$

where $N_{gen}(Z)$ and $N_{rec}(Z)$ are respectively the number of events generated and reconstructed with at least one Z boson. The selection efficiency for events with a Z boson is evaluated to be 0.321.

The values of the selection and jet reconstruction efficiencies involved in these unfolding calculations are summarized in the table 6.7.

6.3.3 Results

Applying the correction described in 6.3.1 and 6.3.2 the following value is obtained for the ratio of the cross sections for the production of events $Z + b\ jet$ and $Z + jet$, with the leading jet in the acceptance region $E_T > 20GeV$, $|\eta| < 1.5$:

$$\frac{\sigma(Z + b\ jet)}{\sigma(Z + jet)} = 0.0202 \pm 0.0044(stat.) \quad (6.15)$$

and for the ratio of the cross sections for the production of events $Z + b\ jet$ and Z , with the leading jet in the acceptance region $E_T > 20GeV$, $|\eta| < 1.5$:

$$\frac{\sigma(Z + b\ jet)}{\sigma(Z)} = 0.00224 \pm 0.00049(stat.) \quad (6.16)$$

In the evaluated ratios of cross sections the efficiencies and systematic uncertainties related to lepton selection cancel exactly. The production cross section $\sigma(Z + b\ jet)$ for events with a Z boson produced in association with at least one jet, and with the leading jet being a $b - jet$ can be calculated multiplying the ratio in 6.16 with the production cross section of Z events $\sigma'(Z)$

(measured or theoretically calculated):

$$\sigma(Z + b\text{ jet}) = \frac{\sigma(Z + b\text{ jet})}{\sigma(Z)} \cdot \sigma'(Z) \quad (6.17)$$

Using the CDF measurement of Z cross section [5], defined to include the leptonic branching fraction

$$\sigma_{meas}(Z, Z \rightarrow l^+l^-) = 254.9 \pm 3.3(stat.) \pm 4.6(sys.) \pm 15.2(lum.) \quad (6.18)$$

the measured cross section for the production of Z boson and at least one jet in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV, with the Z decaying in lepton pair and a b leading jet is:

$$\sigma(Z + b\text{ jet}, Z \rightarrow l^+l^-) = 0.57 \pm 0.12(stat.)pb \quad (6.19)$$

6.4 Systematic uncertainties

The sources of systematic uncertainties relevant for the fitting and unfolding procedure can be grouped in six categories: jet energy scale, correction functions uncertainties, templates selection, background normalization, identification efficiency and unfolding. For every considered systematic effect the measurement is evaluated by the same procedure described in sections 6.2 and 6.3, and the systematic uncertainty has been quoted as the variation range of the measured parameters.

The indetermination on the shape of the jet E_T spectra has an impact on the template shape, on the uncertainty of the correction functions, and also on the jet unfolding correction. In figure 6.8 is presented the E_T distribution of the identifiable leading jet in events with a reconstructed Z boson, compared with the corresponding Alpgen MC prediction. Data and MC simulation show a good agreement except for small values of the leading jet E_T , where non-perturbative effects are more relevant. In figure 6.9 is reported the distribution of the data/MC ratio of the E_T distributions for identifiable leading jets. A fit to the ratio distribution has been performed with both a constant function

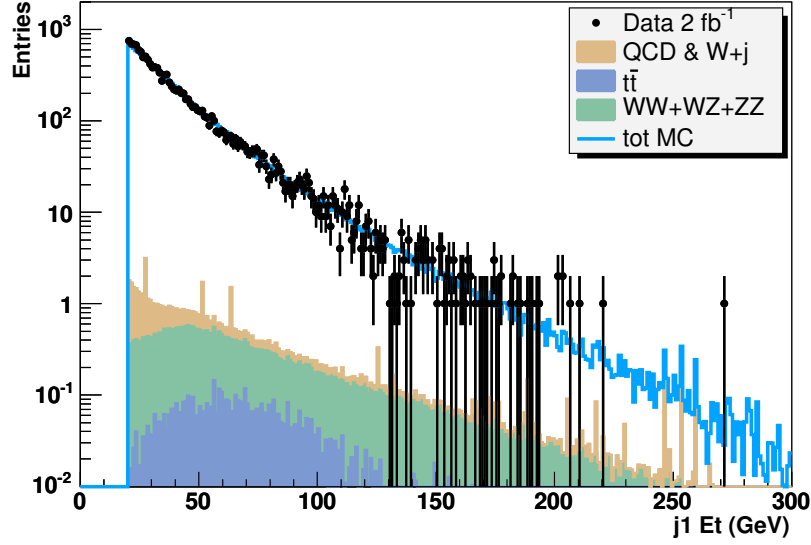


Figure 6.8: E_T distribution of identifiable leading jet in events with a reconstructed Z boson.

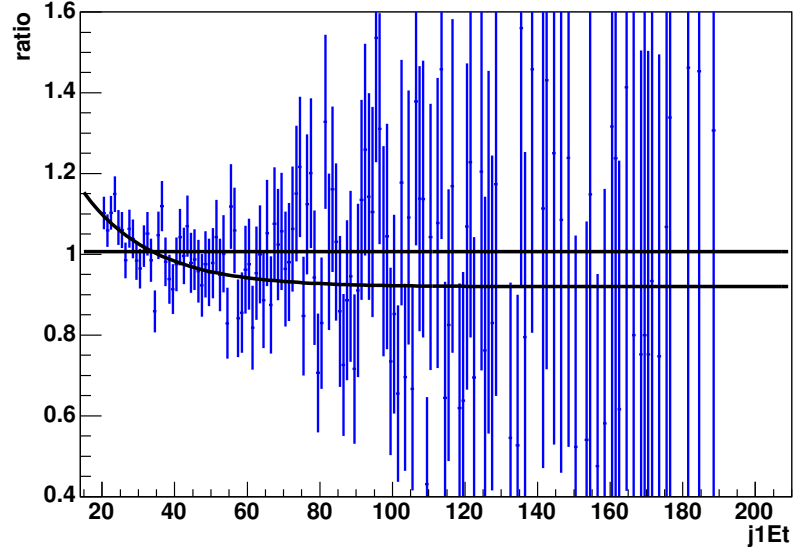


Figure 6.9: Data/MC ratio of the E_T distributions for identifiable leading jet in events with a reconstructed Z boson. The Alpgen MC simulation underestimates the data distribution in the $[20, 40]$ GeV range. The superimposed functions are the results of fit with a constant and an exponential function respectively. (Note the different E_T range respect to figure 6.8.)

(corresponding to a perfect knowledge of the E_T distribution) and an exponential function. The variation between these two functions has been adopted as indetermination on the E_T distribution, and so to evaluate the related systematic uncertainties, the events have been accordingly re-weighted.

6.4.1 Jet energy scale

The indetermination on jet energy scale (“JES”) has an impact on the number of selected events both in data and in MC. In fact a modification of the correction factors for the jet energy implies a migration of events through the selection threshold, with a resulting change in the data spectra and in MC templates. This indetermination does not affect the Neural Network tagger output for a given event, because the algorithm uses the uncorrected value of the transverse energy information.

To evaluate the impact of the jet energy scale uncertainty, the jet energy correction factors are varied of $\pm 1\sigma$, according to the procedure explained in [29], and the measurement has been repeated. The selected data sample have then been fitted with the modified templates for both signal and background.

6.4.2 Correction function uncertainties

The uncertainty related to the calibration procedure can be divided in three categories: the indetermination related to the statistics of the calibration samples, the one due to the selection of the enriched di-jet samples and the one due to the indetermination of the E_T distribution of the calibration samples.

Given the structure of the calibration procedure of the Neural Network templates, the three correction function extracted are correlated. In section 4.4 a toy MC method to produce triplets of correction functions properly correlated has been described. A set of ~ 1000 triplets of correction functions has been applied to standard templates, repeating the fraction fit to data spectra every time. The RMS of the distribution of the measured fractions has been taken as systematic uncertainty due to the indetermination of the correction functions.

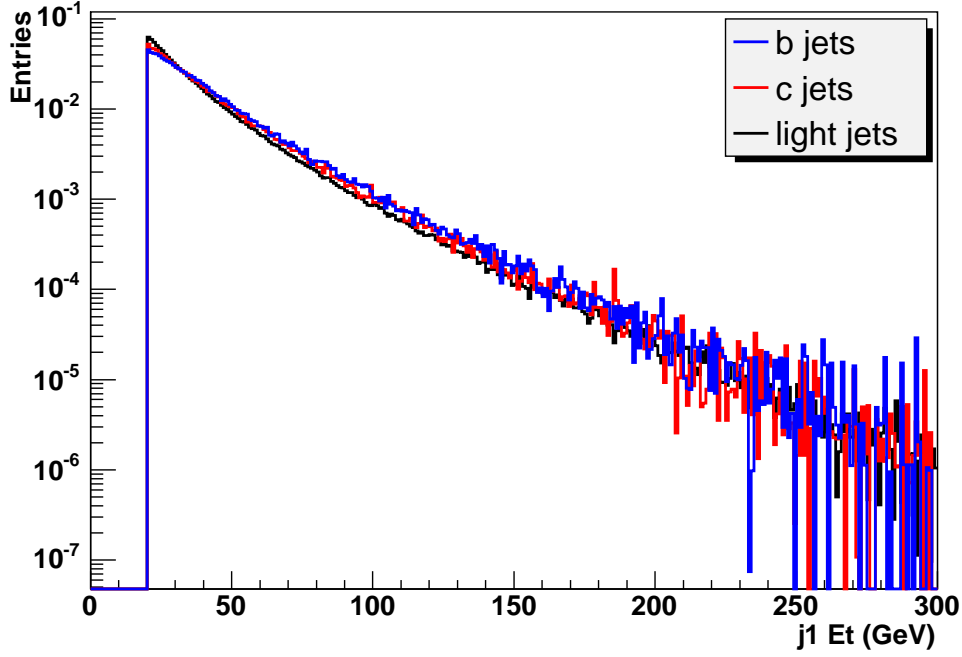


Figure 6.10: Comparison of the b , c and $light - jet$ E_T spectra predicted by the Alpgen MC for $Z + jets$ events.

To evaluate the systematic uncertainty due to the definition of the calibration samples, the B and C enriched samples have been altered changing the value of the cut in $LBLnn$ from 0.6 to 0.5 and 0.7. Full calibrations have been performed with both the modified samples, and the corresponding triplets of extracted correction functions have been applied to the fraction fit on data.

The indetermination on the E_T distribution affects also the correction functions because of the re-weight procedure of the calibration samples described in section 6.1.3. In particular the $b - jet$ and $c - jet$ E_T distributions, can not be extracted from data sample, and the re-weight procedure for the Heavy Flavour components relies only on the MC prediction. The comparison of the E_T distributions predicted by Alpgen for b , c and $light - jets$ is reported in figure 6.10. To evaluate the systematic uncertainty related to the indetermination of the $b - jet$ and $c - jet$ E_T distributions, a complete set of correction functions is determined re-weighting all the calibration samples to the E_T dis-

tribution predicted by Alpgen for $Z + \text{light} - \text{jet}$ events. The variation range of the measured fractions obtained substituting this set of correction functions to the nominal one, is taken as systematic uncertainty.

6.4.3 Uncertainty on templates

Possible differences in E_T distribution between data and MC simulation are accounted for re-weighting the MC distributions with the exponential function described in section 6.4.2.

Furthermore, since it is not *a priori* known whether the jets contain one or two heavy quarks, a systematic uncertainty is estimated varying in the b and c templates respectively the fraction of double b and double c events from 0 to 3 times the value predicted by the Alpgen MC, as shown in figure 6.11.

6.4.4 Background normalization

The uncertainty on the normalization of the background due to fake electrons is estimated to be $\pm 50\%$, as described in section 5.2.5. The same uncertainty has been applied to the fake muon background, due the little statistic of this kind of events.

The uncertainties on the normalization of the backgrounds due to di-boson and $t\bar{t}$ are estimated to be $\pm 20\%$. These uncertainties include the NLO cross section uncertainty estimated by varying the factorization scale by a factor two ([28]), the cross section uncertainty due to the uncertainty on the top quark mass, experimental uncertainties (e.g. the lepton identification and b jet tagging efficiencies), and the uncertainty on the identification efficiency scale factor.

In the quoted systematic uncertainties is included also a 6% uncertainty on the measurement of the integrated luminosity [5].

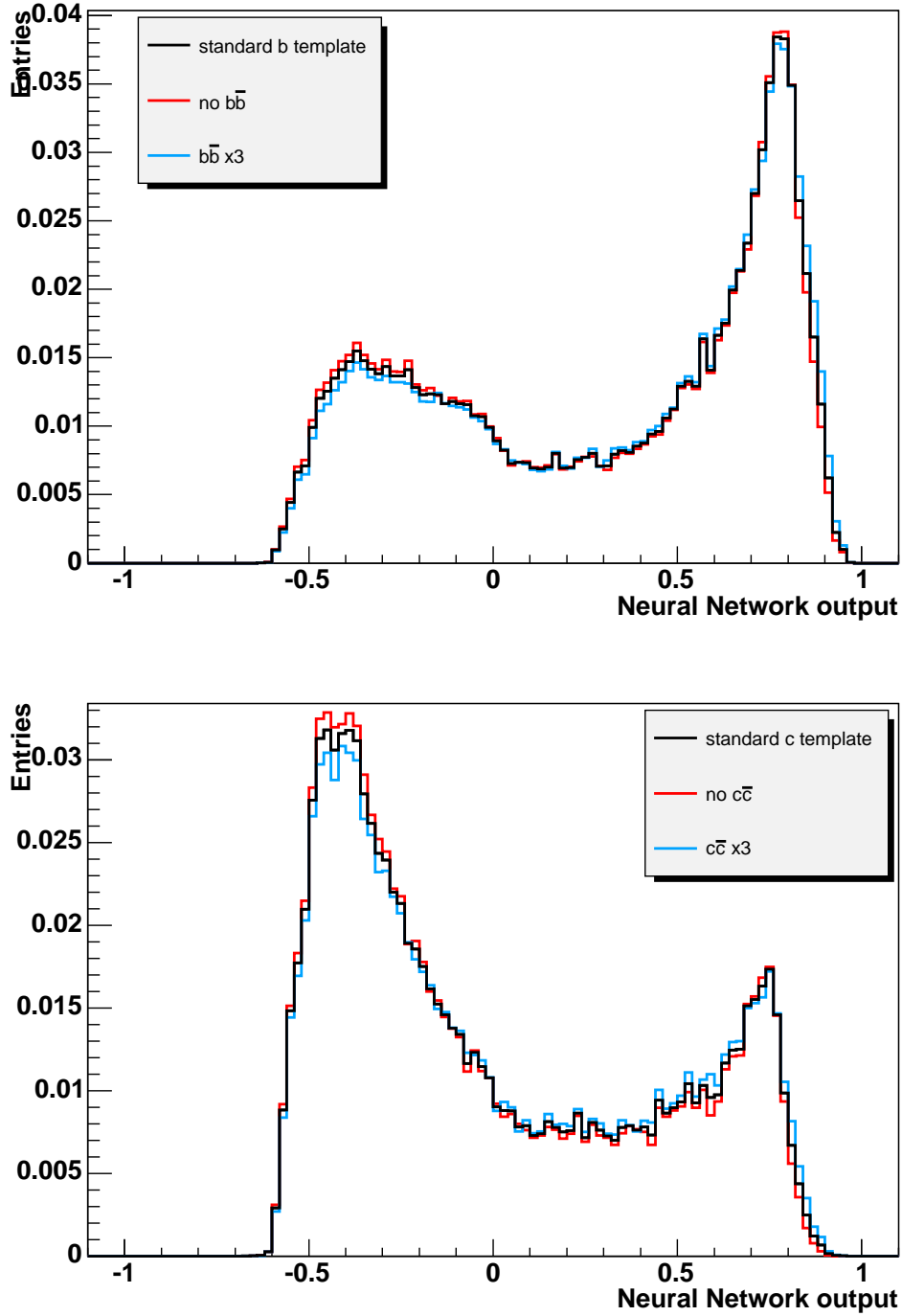


Figure 6.11: Comparison of the b (upper panel) and c (lower panel) templates when the fraction of double b and double c events, respectively, is varied from 0 to 3 times the value predicted by the Alpgen MC.

Systematic uncertainty	$Z + b\text{ jet}$ [%]	$Z + c\text{ jet}$ [%]	$Z + \text{light jet}$ [%]
JES	1.6	4.5	0.02
cf uncertainty	10	27	0.8
cf enriching	5.5	19	0.22
cf E_T re-weight	3.9	16	0.1
$b/b\bar{b}$, $c/c\bar{c}$	10	30	0.1
E_T re-weight	0.8	1.0	0.03
Fake	2.4	2.5	0.015
Top	0.8	0.5	0.005
Di-Boson	-	1.0	0.01
Identification efficiency	1.8	2.2	0.05
Jet unfolding	0.2	0.01	-
Total	16	48	0.86

Table 6.8: Summary of the relative systematics uncertainties.

6.4.5 Identification efficiency uncertainty

The identification efficiency correction, relies on the independence of the identification efficiency scale factor from the jet flavour. To account for possible deviations from this hypothesis, an uncertainty equal to $\pm 2\%$ for the identification efficiency is considered independently for each flavour, and the maximum variation range for each fitted fraction is assumed as systematic uncertainty.

6.4.6 Jet unfolding uncertainty

The impact of the E_T uncertainty on the jet reconstruction efficiencies is evaluated applying the E_T re-weight described in section 6.4 to the MC distributions both at calorimeter and at particle level. The variation range of the unfolding ratios is then assumed as systematic uncertainty.

	CDF data	NLO prediction	Alpgen prediction
$\frac{\sigma(Z+b\,jet)}{\sigma(Z+jet)}$	$0.0202 \pm 0.0044 \pm 0.0032$	0.017	0.0121
$\frac{\sigma(Z+b\,jet)}{\sigma(Z)}$	$0.00224 \pm 0.00049 \pm 0.00036$	0.0021	0.0013
$\sigma(Z + b\,jet)$	$0.57 \pm 0.12 \pm 0.09\,pb$	$0.54\,pb$	$0.34\,pb$

Table 6.9: Measured cross section and ratio of cross sections for $Z + b\,jet$ events, compared with NLO and Alpgen prediction. The cross section value predicted by Alpgen is calculated with the same ratio technique used for data. The NLO prediction is evaluated by a MCFM ([13], [14], [15]) simulation. The measured values are reported with their statistical and systematic errors.

6.4.7 Summary of systematic uncertainties

In table 6.8 all the systematic uncertainties evaluated for this measurement are summarized. The most relevant systematic uncertainties for the $Z + b\,jet$ fraction are the Neural Network shape calibration uncertainty and the indetermination on the content of double b and double c in the leading jet.

6.5 Results and future perspectives

Including the systematic uncertainties, the measured value for the ratio of the cross sections for the production of events $Z + b\,jet$ and $Z + jet$, with the leading jet in the acceptance region $E_T > 20GeV$, $|\eta| < 1.5$ is:

$$\frac{\sigma(Z + b\,jet)}{\sigma(Z + jet)} = 0.0202 \pm 0.0044(stat.) \pm 0.0032(sys.) \quad (6.20)$$

and the measured value for the production cross section for the associated production of a Z boson and at least one jet, with the leading jet being a $b - jet$ is:

$$\sigma(Z + b\,jet, Z \rightarrow l^+l^-) = 0.57 \pm 0.12(stat.) \pm 0.09(sys.)pb \quad (6.21)$$

In table 6.9 the measured cross section is compared with corresponding MC predictions. The NLO calculation is performed by the MCFM ([13], [14], [15]) program, and is not corrected for underlying event and hadronization.

The technique presented in the previous sections provides a measurement of the production of $Z+b$ events with a 20% statistical error and a 16% systematic error. The increasing statistics delivered by Tevatron will make possible a significant reduction of both the statistical and systematic uncertainties on the measurement. In fact an increased statistics can be widely exploited also to reduce the uncertainty on the Neural Network tagger calibration, that is one of the main contributions to the systematic error. The other main source of systematic error is the uncertainty on *double* – b and *double* – c content in identifiable jets. This uncertainty can be significantly reduced by specific studies on the double Heavy Flavour fraction of identifiable jets. In this context also the finalization of the development of the five-component tagger, with its corresponding calibration, will be particularly useful.

Increasing the statistics (with the consequent reduction of the Neural Network tagger calibration uncertainty) and improving the fitting procedure used, it will be possible to provide also a measurement of the production cross section of $Z + c - jets$ events.

Furthermore an increased statistics will make accessible the measurement of differential distributions of the leading jet in $Z + HF - jets$ events, and the measurement of the Heavy Flavour content in events of associated production of a Z boson and two or more jets.

The presented technique is quite general and could be easily adapted to similar high- P_T analysis, like $W + b$.

The information provided by the $Z + b$ measurement is a decisive element in the search of the Higgs boson in the mass range accessible to Tevatron. Moreover the $Z + b$ and $Z + c$ production cross sections can provide important information for the parton *pdf* fit.

Appendix A

Fraction fits

The fraction and *correction function* fits are based on the minimization performed by Minuit of the following Negative Log Likelihood function obtained assuming Poisson distribution of the bin content:

$$-\ln \mathcal{L} = -\sum_{k=1}^n \left[d_k \ln \left(\sum_{j=1}^M f_j \phi_{jk} \frac{t_{jk}}{N_j} \right) \right] \quad (\text{A.1})$$

where n is the bins number, M is the number of flavors, d_k is the content of the k -th bin of the spectra to be fitted, f_j is the fraction of the j -th flavor component, t_{jk} is the content of the k -th bin in the j -th MC template and N_j is the total number of entries in the j -th MC template.

The weight ϕ_{jk} of the j -th flavor component in the k -th bin is obtained evaluating the *correction function* in the bin center (x_k).

Corrected templates are constrained to conserve their integral through the iterations and then to be equal to initial integral:

$$\sum_{k=0}^N \phi_{jk} t_{jk} = \sum_{k=0}^N t_{jk} \quad (\text{A.2})$$

This constraint is implemented in the Likelihood minimization.

In the present analysis the functional form of the correction functions is assumed to be linear

$$\phi_{jk} = p_{0,j} + p_{1,j} \cdot x_k \quad (\text{A.3})$$

even if the fit procedure can be applied to a wider range of functions. As

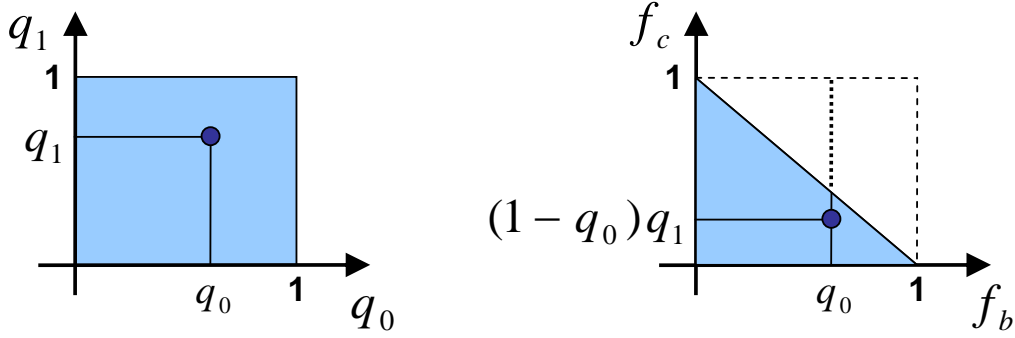


Figure A.1: Mapping scheme used in the three-component fraction fit: the cyan square region in the left panel corresponds to the triangular region, allowed for the independent fractions, shown in the right panel.

a consequence in each template all the weights depend only on one parameter. In the *correction function* fit, fractions are fixed and $p_{0,j}$ is the only free parameter.

In the fraction fit the *correction functions* (the weights) are fixed, the constraint $f_b + f_c + f_l = 1$ and the following mapping scheme, over the last constraint, is implemented:

$$f_b = q_0 \tag{A.4}$$

$$f_c = (1 - q_0) \cdot q_1 \tag{A.5}$$

$$f_l = (1 - q_0) \cdot (1 - q_1) \tag{A.6}$$

where bounding q_0 and q_1 in $[0, 1]$ corresponds in bounding values for the fractions in the allowed physical region, as shown in figure A.1.

Bibliography

- [1] S. L. Glashow, “Partial Symmetry Of Weak Interactions”, *Nucl. Phys.*, 22:579-588, 1961.
- [2] A. Salam, “Weak And Electromagnetic Interactions”, Originally printed in *Svartholm: Elementary Particle Theory, Proceedings Of The Nobel Symposium Held 1968 At Lerum, Sweden*, Stockholm 1968, 367-377.
- [3] S. Weinberg, “Implications Of Dynamical Symmetry Breaking”, *Phys. Rev.*, D13:974-996, 1976.
- [4] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. Nadolsky and W. K. Tung, “New generation of parton distributions with uncertainties from global QCD analysis,” JHEP **0207** (2002) 012 [arXiv:hep-ph/0201195].
- [5] D. E. Acosta *et al.* [CDF II Collaboration], “First measurements of inclusive W and Z cross sections from Run II of the Tevatron collider,” *Phys. Rev. Lett.* **94** (2005) 091803 [arXiv:hep-ex/0406078].
- [6] P. J. Sutton, A. D. Martin, R. G. Roberts and W. J. Stirling, *Phys. Rev.* **D45**, 2349 (1992).
- [7] P. J. Rijken and W. L. van Neerven, *Phys. Rev.* **D51**, 44 (1995).
- [8] R. Hamberg, W. L. van Neerven and W. B. Kilgore, *Nucl. Phys.* **B359**, 343 (1991).
- [9] R. V. Harlander and W. B. Kilgore, *Phys. Rev. Lett.* **88**, 201801 (2002).

-
- [10] E. Accomando *et al.*, Physics Reports **299** 1-78 (1998)
- [11] S. Dawson, C. B. Jackson, L. Reina and D. Wackerth, Phys. Rev. Lett. **94**, (2005) 031802 [arXiv:hep-ph/0408077]
- [12] J. Campbell et al., in “Les Houches 2003: Physics at TeV Colliders”, [arXiv:hep-ph/0405302]
- [13] J. Campbell, R. K. Ellis, F. Maltoni and S. Willenbrock, “Production of a Z boson and two jets with one heavy-quark tag,” Phys. Rev. D **73** (2006) 054007 [Erratum-ibid. D **77** (2008) 019903] [arXiv:hep-ph/0510362].
- [14] F. Maltoni, T. McElmurry and S. Willenbrock, “Inclusive production of a Higgs or Z boson in association with heavy quarks,” Phys. Rev. D **72** (2005) 074024 [arXiv:hep-ph/0505014].
- [15] J. Campbell, R. K. Ellis, F. Maltoni and S. Willenbrock, “Associated production of a Z boson and a single heavy-quark jet,” Phys. Rev. D **69** (2004) 074021 [arXiv:hep-ph/0312024].
- [16] G. C. Blazey *et al.*, “Run II jet physics,” arXiv:hep-ex/0005012.
- [17] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A. D. Polosa, “ALPGEN, a generator for hard multiparton processes in hadronic collisions,” JHEP **0307** (2003) 001 [arXiv:hep-ph/0206293].
- [18] F. Caravaglios and M. Moretti, “An algorithm to compute Born scattering amplitudes without Feynman graphs”, *Phys. Lett.*, B358:332-338, 1995.
- [19] F. Caravaglios, M. Mangano, Michelangelo, M. Moretti and R. Pittau, “A new approach to multi-jet calculations in hadron collisions”, *Nucl. Phys.*, B539:215-232, 1999.
- [20] T. Sjostrand, S. Mrenna and P. Skands, “PYTHIA 6.4 physics and manual,” JHEP **0605** (2006) 026 [arXiv:hep-ph/0603175].

- [21] G. Corcella *et al.*, “HERWIG 6: An event generator for hadron emission reactions with interfering gluons (including supersymmetric processes),” JHEP **0101** (2001) 010 [arXiv:hep-ph/0011363].
- [22] A. Artikov, M. Giunta *et al.* “CDF Central Preshower and Crack Detector Upgrade”, submitted to Nucl. Inst. and Meth. A in February 2007.
- [23] A.A.V.V. *The CDF II Detector Technical Design Report*, FERMILAB-Pub-96/390-E, 1996.
- [24] D. Jeans “B tagging at CDF”, Proceedings for HCP2005 conference, CDF public note 7854.
- [25] <http://www.phi-t.de/>
- [26] R. Brun, F. Rademakers, P. Canal, I. Antcheva, D. Buskulic, “ROOT. An Object-Oriented Data Analysis Framework” <http://root.cern.ch/>
- [27] J. Marriner, *Secondary Vertex Fit with Mass and Pointing Constraints (CTVMFT)* Tech. Rep.CDF/DOC/SEC_VTX/PUBLIC/1996, 1993.
- [28] Cacciari *et al.*, JHEP **0404** (2004) 068
- [29] A. Bhatti *et al.*, “Determination of the jet energy scale at the Collider Detector at Fermilab,” Nucl. Instrum. Meth. A **566** (2006) 375 [arXiv:hep-ex/0510047].